INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA

Álgebra Linear

Coleção Matemática Universitária

Elon Lages Lima

Álgebra Linear

Este livro ganhou o prêmio Jabuti de Ciências Exatas e Tecnologia, outorgado pela Câmara Brasileira do Livro, em 1996

COLEÇÃO MATEMÁTICA UNIVERSITÁRIA

Álgebra Linear

Elon Lages Lima

INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA

Impresso no Brasil / Printed in Brazil

Capa: Rodolfo Capeto, Noni Geiger e Sérgio R. Vaz

Títulos Publicados:

- Análise Real, vol. 1: Funções de uma Variável – Elon Lages Lima
- EDP. Um Curso de Graduação – Valéria Iório
- Curso de Álgebra, Volume 1 – Abramo Hefez
- Álgebra Linear – Elon Lages Lima
- Introdução às Curvas Algébricas Planas – Israel Vainsencher
- Equações Diferenciais Aplicadas – Djairo G. de Figueiredo e Aloisio Freiria Neves
- Geometria Diferencial – Paulo Ventura Araújo
- Introdução à Teoria dos Números – José Plínio de Oliveira Santos
- Cálculo em uma Variável Complexa – Marcio G. Soares
- Geometria Analítica e Álgebra Linear – Elon Lages Lima
- Números Primos: Mistérios e Recordes – Paulo Ribenboim
- Análise no Espaço R^n – Elon Lages Lima
- Análise Real, vol. 2: Funções de n Variáveis – Elon Lages Lima
- Álgebra Exterior – Elon Lages Lima
- Equações Diferenciais Ordinárias – Claus Ivo Doering e Artur Oscar Lopes
- Análise Real, vol. 3: Análise Vetorial – Elon Lages Lima
- Álgebra Linear. Exercícios e Soluções – Ralph Costa Teixeira
- Números Primos. Velhos Mistérios e Novos Recordes – Paulo Ribenboim

Distribuição:
IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
e-mail: ddic@impa.br
http://www.impa.br

Prefácio

Álgebra Linear é o estudo dos espaços vetoriais e das transformações lineares entre eles. Quando os espaços têm dimensões finitas, as transformações lineares possuem matrizes. Também possuem matrizes as formas bilineares e, mais particularmente, as formas quadráticas. Assim a Álgebra Linear, além de vetores e transformações lineares, lida também com matrizes e formas quadráticas. São numerosas e bastante variadas as situações, em Matemática e em suas aplicações, onde esses objetos ocorrem. Daí a importância central da Álgebra Linear no ensino da Matemática.

O presente livro apresenta uma exposição introdutória de Álgebra Linear. Ele não pressupõe conhecimentos anteriores sobre o assunto. Entretanto convém lembrar que a posição natural de um tal curso no currículo universitário vem após um semestre (pelo menos) de Geometria Analítica a duas e três dimensões, durante o qual o estudante deve adquirir alguma familiaridade, em nível elementar, com a representação algébrica de idéias geométricas e vice-versa.

Tornou-se quase obrigatório, já faz alguns anos, dedicar as primeiras sessenta ou mais páginas de todo livro de Álgebra Linear ao estudo dos sistemas de equações lineares pelo método da eliminação gaussiana, motivando assim a introdução das matrizes e dos determinantes. Somente depois disso são definidos os espaços vetoriais.

Esse costume não é seguido neste livro, cuja primeira sentença é a definição de espaço vetorial. Mencionarei três razões para isso: (a) A definição de Álgebra Linear dada acima; (b) Não vejo vantagem em longas motivações; (c) Sistemas lineares são entendidos mais inteligentemente depois que já se conhecem os conceitos básicos de Álgebra Linear. De resto, esses conceitos (núcleo, imagem, base, posto, subespaço, etc), quando estudados independentemente, têm muitas outras aplicações.

O método da eliminação gaussiana é apresentado na Seção 9 e retomado na Seção 17. Ele é aplicado para obter respostas a vários outros problemas além da resolução de sistemas lineares.

O livro é dividido em vinte e duas seções. As oito primeiras desenvolvem os conceitos fundamentais e as proposições básicas, que formam a linguagem mínima necessária para falar inteligentemente sobre Álgebra Linear. A nona seção faz a primeira aplicação dessas idéias, tratando da eliminação gaussiana.

A partir da Seção 10, os espaços dispõem de produto interno, o que possibilita o emprego de evocativas noções geométricas como perpendicularismo, comprimento, distância, etc. São destacados tipos particulares de operadores lineares, cujas propriedades especiais são demonstradas nas Seções 13, 14 e 15. O Teorema Espectral para operadores auto-adjuntos é provado na Seção 13, onde se demonstra também o Teorema dos Valores Singulares (Teorema 13.10), cuja grande utilidade não corresponde à sua conspícua ausência na maioria dos textos elementares.

Outro assunto igualmente importante e igualmente esquecido no ensino da Álgebra Linear é a pseudo-inversa, que expomos na Seção 16. Trata-se de um tópico fácil, atraente, de grande apelo geométrico, que constitui um bom campo de aplicação para os conceitos anteriormente estudados.

A Seção 17 é um interlúdio matricial, onde se mostra como as propriedades das transformações lineares estudadas antes se traduzem imediatamente em fatos não-triviais sobre matrizes, principalmente algumas decomposições de grande utilidade nas computações.

As formas bilineares e quadráticas são estudadas na Seção 18, onde é estabelecida a correspondência fundamental (isomorfismo) entre formas e operadores (Teorema 18.2) e provado o Teorema dos Eixos Principais (Teorema 18.3), que é a versão do Teorema Espectral para formas quadráticas. É ainda exposto o método de Lagrange para reduzir uma forma quadrática a uma soma (ou diferença) de quadrados e é feito um estudo das superfícies quádricas.

Os determinantes são estudados na Seção 19, onde se define diretamente o determinante de um operador sem recurso a bases nem matrizes. Em seguida, o determinante de uma matriz $n \times n$ é caracterizado como a única função n-linear alternada de suas colunas (ou linhas) que assume o valor 1 na matriz unitária. A colocação dos

determinantes quase no final do livro, depois de já terem sido estabelecidos os resultados principais da Álgebra Linear e ensinados os métodos mais eficientes para resolver sistemas, inverter matrizes etc, é uma atitude deliberada, que visa pôr esse conceito em seu devido lugar. Trata-se de uma noção de grande importância teórica, indispensável em várias áreas da Matemática, a qual foi, e ainda não deixou inteiramente de ser, equivocadamente considerada como instrumento computacional. Usar a Regra de Cramer para resolver um sistema linear, ou calcular o determinante de um operador para ver se ele é invertível ou não, são métodos que funcionam bem no caso 2×2, e até mesmo 3×3, mas se tornam altamente inviáveis a partir daí.

Depois que se têm os determinantes, o polinômio característico é estudado na Seção 20. Esse estudo se completa na Seção 21 com a introdução dos espaços vetoriais complexos, nos quais vale o notável fato de que todo operador possui autovetores, logo pode ser triangularizado. Este resultado é devidamente explorado, o que concede a esta seção um ar de happy ending para a teoria, mas não o fim do livro.

A seção final, número 22, apresenta uma breve exposição das equações a diferenças finitas, essencialmente limitada às equações (e sistemas) lineares de segunda ordem. Basicamente, trata-se de obter métodos eficazes de calcular as potências sucessivas de um operador ou de suas matrizes.

Esta introdução à Álgebra Linear reflete uma longa experiência como usuário do assunto e, nos últimos dez anos, como professor. Ao escrevê-la, fui influenciado pelas reações dos meus alunos, suas participações nas aulas e suas palavras de incentivo. Um agradecimento especial por esse motivo é devido aos estudantes da E.P.G.E. da Fundação Getúlio Vargas. Agradeço ao meu colega Jonas de Miranda Gomes por me ter convencido de que ainda havia lugar para mais um livro nesta área e por suas sugestões, sempre objetivas, que contribuíram para melhorar a comunicabilidade. Agradeço também a Wilson L. de Góes pela incrível eficiência e grande boa vontade na preparação do manuscrito.

Rio de Janeiro, maio de 1995
Elon Lages Lima

Prefácio da Segunda Edição

A boa acolhida dispensada à primeira edição, esgotada rapidamente, animou-me a fazer nesta algumas modificações, que enumero a seguir.

Foi feita uma extensa revisão do texto, eliminando-se vários erros de impressão, exercícios incorretamente propostos e trechos obscuros ou imprecisos. Para este trabalho, vali-me da colaboração de diversos leitores, dentre os quais destaco, de modo muito especial, o Professor Florêncio Guimarães, que elaborou uma lista minuciosa de correções. A todos esses amigos registro meus sinceros agradecimentos.

O número de exercícios foi consideravelmente aumentado com a inclusão, em especial, de mais problemas elementares de natureza computacional, visando fazer com que os leitores menos experientes ganhem confiança em si ao lidarem com assuntos novos.

A Seção 15 foi inteiramente reescrita, passando a tratar dos operadores normais em espaços vetoriais reais, um assunto fácil, atraente e muitas vezes negligenciado. A antiga Seção 15 (operadores anti-simétricos) tornou-se um mero caso particular. Sem esforço (nem espaço) adicional, o tratamento ganhou uma abrangência bem maior.

Atendendo a vários pedidos, acrescentei ao livro um Apêndice sobre a forma canônica de Jordan, tratando esse tema de modo simples, não apenas sob o ponto de vista matricial mas formulando-o também sob o aspecto de operadores.

Rio de Janeiro, setembro de 1996
Elon Lages Lima

Prefácio da Oitava Edição

Esta edição, além de conter novas correções sugeridas pela atenta vigilância do Professor Florêncio Guimarães, deu-me oportunidade de acrescentar à lista de indicações bibliográficas o livro do Professor Ralph Costa Teixeira, que traz as soluções de todos os exercícios aqui propostos.

Rio de Janeiro, outubro de 2009
Elon Lages Lima

Conteúdo

1

Espaços Vetoriais

A noção de espaço vetorial é a base do estudo que faremos; é o terreno onde se desenvolve toda a Álgebra Linear. Esta seção apresenta os axiomas de espaço vetorial, deduz suas conseqüências mais imediatas e exibe os exemplos mais importantes dessa noção.

Um *espaço vetorial* E é um conjunto, cujos elementos são chamados *vetores*, no qual estão definidas duas operações: a *adição*, que a cada par de vetores $u, v \in E$ faz corresponder um novo vetor $u+v \in E$, chamado a *soma* de u e v, e a *multiplicação por um número real*, que a cada número $\alpha \in \mathbb{R}$ e a cada vetor $v \in E$ faz corresponder um vetor $\alpha \cdot v$, ou αv, chamado o *produto* de α por v. Essas operações devem satisfazer, para quaisquer $\alpha, \beta \in \mathbb{R}$ e $u, v, w \in E$, as condições abaixo, chamadas os *axiomas* de espaço vetorial:

comutatividade: $u + v = v + u$;

associatividade: $(u + v) + w = u + (v + w)$ e $(\alpha\beta)v = \alpha(\beta v)$;

vetor nulo: existe um vetor $0 \in E$, chamado *vetor nulo*, ou *vetor zero*, tal que $v + 0 = 0 + v = v$ para todo $v \in E$;

inverso aditivo: para cada vetor $v \in E$ existe um vetor $-v \in E$, chamado o *inverso aditivo*, ou o *simétrico* de v, tal que $-v + v = v + (-v) = 0$;

distributividade: $(\alpha + \beta)v = \alpha v + \beta v$ e $\alpha(u + v) = \alpha u + \alpha v$;

multiplicação por 1: $1 \cdot v = v$.

Observação: O mesmo símbolo 0 representa o vetor nulo e o número zero.

Exemplo 1.1. Para todo número natural n, o símbolo $\mathbb{R}^n$ representa o *espaço vetorial euclidiano* n-*dimensional*. Os elementos de $\mathbb{R}^n$ são as listas ordenadas $u = (\alpha_1, \ldots, \alpha_n)$, $v = (\beta_1, \ldots, \beta_n)$ de números reais.

Por definição, a igualdade vetorial $u = v$ significa as n igualdades numéricas $\alpha_1 = \beta_1, \ldots, \alpha_n = \beta_n$.

Os números $\alpha_1, \ldots, \alpha_n$ são chamados as *coordenadas* do vetor u. As operações do espaço vetorial $\mathbb{R}^n$ são definidas pondo

$$u + v = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n),$$
$$\alpha \cdot u = (\alpha\alpha_1, \ldots, \alpha\alpha_n).$$

O *vetor zero* é, por definição, aquele cujas coordenadas são todas iguais a zero: $0 = (0, 0, \ldots, 0)$.

O inverso aditivo de $u = (\alpha_1, \ldots, \alpha_n)$ é $-u = (-\alpha_1, \ldots, -\alpha_n)$. Verifica-se, sem dificuldade, que estas definições fazem de $\mathbb{R}^n$ um espaço vetorial. Para $n = 1$, tem-se $\mathbb{R}^1 = \mathbb{R}$ = reta numérica. $\mathbb{R}^2$ é o plano euclidiano e $\mathbb{R}^3$ é o espaço euclidiano tridimensional da nossa experiência cotidiana.

Para ajudar a compreensão, os vetores de $\mathbb{R}^2$ e $\mathbb{R}^3$ podem ser representados por flechas com origem no mesmo ponto O. A soma $u+v$ é a flecha que liga a origem O ao vértice que lhe é oposto no paralelogramo que tem u e v como lados. (Veja Figura 1.1.)

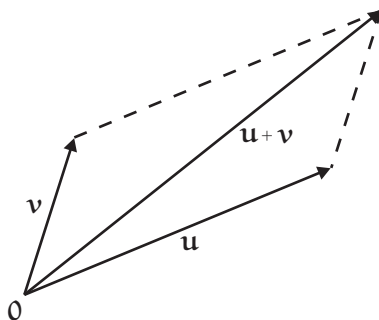



Figura 1.1 – Soma de vetores.

Por sua vez, o produto αu é a flecha colinear a u, de comprimento α vezes o comprimento de u, com o mesmo sentido de u se $\alpha > 0$ e com sentido oposto se $\alpha < 0$.

Exemplo 1.2. Os elementos do espaço vetorial $\mathbb{R}^\infty$ são as seqüências infinitas $u = (\alpha_1, \ldots, \alpha_n, \ldots)$, $v = (\beta_1, \ldots, \beta_n, \ldots)$ de números reais. O elemento zero de $\mathbb{R}^\infty$ é a seqüência $0 = (0, \ldots, 0, \ldots)$, formada por infinitos zeros, e o inverso aditivo da seqüência $u = (\alpha_1, \ldots, \alpha_n, \ldots)$ é $-u = (-\alpha_1, \ldots, -\alpha_n, \ldots)$. As operações de adição e multiplicação por um número real são definidas por

$$u + v = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n, \ldots),$$
$$\alpha \cdot u = (\alpha\alpha_1, \ldots, \alpha\alpha_n, \ldots).$$

Exemplo 1.3. Uma *matriz* (real) $m \times n$ $\mathbf{a} = [a_{ij}]$ é uma lista de números reais a_{ij} com índices duplos, onde $1 \leq i \leq m$ e $1 \leq j \leq n$. Costuma-se representar a matriz $\mathbf{a}$ como um quadro numérico com m linhas e n colunas, no qual o elemento a_{ij} situa-se no cruzamento da i-ésima linha com a j-ésima coluna:

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

O vetor $(a_{i1}, a_{i2}, \ldots, a_{in}) \in \mathbb{R}^n$ é o i-ésimo *vetor-linha* da matriz $\mathbf{a}$ e o vetor $(a_{1j}, a_{2j}, \ldots, a_{mj}) \in \mathbb{R}^m$ é o j-ésimo *vetor-coluna* de $\mathbf{a}$. Quando $m = n$, diz-se que $\mathbf{a}$ é uma *matriz quadrada*. O conjunto $M(m \times n)$ de todas as matrizes $m \times n$ torna-se um espaço vetorial quando nele se define a soma das matrizes $\mathbf{a} = [a_{ij}]$ e $\mathbf{b} = [b_{ij}]$ como $\mathbf{a} + \mathbf{b} = [a_{ij} + b_{ij}]$ e o produto da matriz $\mathbf{a}$ pelo número real α como $\alpha\mathbf{a} = [\alpha a_{ij}]$. A matriz nula $\mathbf{0} \in M(m \times n)$ é aquela formada por zeros e o inverso aditivo da matriz $\mathbf{a} = [a_{ij}]$ é $-\mathbf{a} = [-a_{ij}]$.

Exemplo 1.4. Seja X um conjunto não-vazio qualquer. O símbolo $\mathcal{F}(X; \mathbb{R})$ representa o conjunto de todas as funções reais $f, g \colon X \to \mathbb{R}$. Ele se torna um espaço vetorial quando se definem a soma $f + g$ de duas funções e o produto $\alpha \cdot f$ do número α pela função f da maneira natural:

$$(f + g)(x) = f(x) + g(x), \quad (\alpha f)(x) = \alpha \cdot f(x).$$

Variando o conjunto X, obtêm-se diversos exemplos de espaços vetoriais da forma $\mathcal{F}(X;\mathbb{R})$. Por exemplo, se $X = \{1,\ldots,n\}$ então $\mathcal{F}(X;\mathbb{R}) = \mathbb{R}^n$; se $X = \mathbb{N}$ então $\mathcal{F}(X;\mathbb{R}) = \mathbb{R}^\infty$; se X é o produto cartesiano dos conjuntos $\{1,\ldots,m\}$ e $\{1,\ldots,n\}$ então $\mathcal{F}(X;\mathbb{R}) = M(m \times n)$.

Outros exemplos de espaços vetoriais ocorrem como subespaços, como veremos a seguir.

Valem num espaço vetorial, como conseqüências dos axiomas, as regras operacionais habitualmente usadas nas manipulações numéricas. Vejamos algumas delas.

1. *Se* $w + u = w + v$ *então* $u = v$. *Em particular,* $w + u = w$ *implica* $u = 0$ *e* $w + u = 0$ *implica* $u = -w$.

Com efeito, da igualdade $w + u = w + v$ segue-se que

$$\begin{aligned} u = 0 + u &= (-w + w) + u \\ &= -w + (w + u) \\ &= -w + (w + v) \\ &= (-w + w) + v \\ &= 0 + v = v. \end{aligned}$$

Em particular, $w + u = w$ implica $w + u = w + 0$, logo $u = 0$. E se $w + u = 0$ então $w + u = w + (-w)$ logo $u = -w$.

2. *Dados* $0 \in \mathbb{R}$ *e* $v \in E$ *tem-se* $0 \cdot v = 0 \in E$. *Analogamente, dados* $\alpha \in \mathbb{R}$ *e* $0 \in E$, *vale* $\alpha \cdot 0 = 0$.

Com efeito, $v + 0 \cdot v = 1 \cdot v + 0 \cdot v = (1+0) \cdot v = 1 \cdot v = v$, logo $0 \cdot v = 0$ como vimos acima. De modo análogo, como $\alpha \cdot 0 + \alpha \cdot 0 = \alpha \cdot (0+0) = \alpha \cdot 0$, segue-se de 1) que $\alpha \cdot 0 = 0$.

3. *Se* $\alpha \neq 0$ *e* $v \neq 0$ *então* $\alpha \cdot v \neq 0$.

Com efeito, se fosse $\alpha \cdot v = 0$ então $v = 1 \cdot v = (\alpha^{-1} \cdot \alpha)v = \alpha^{-1} \cdot (\alpha v) = \alpha^{-1} \cdot 0 = 0$, isto é, teríamos $v = 0$.

4. $(-1) \cdot v = -v$.

Com efeito,

$$v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1 + (-1)) \cdot v = 0 \cdot v = 0,$$

logo $(-1)v = -v$, pela regra 1.

No que se segue, escreveremos $u - v$ para significar $u + (-v)$. Evidentemente,

$$u - v = w \Leftrightarrow u = v + w.$$

Exemplo 1.5. Sejam $u = (a, b)$ e $v = (c, d)$ vetores em $\mathbb{R}^2$ com $u \neq 0$, isto é, $a \neq 0$ ou $b \neq 0$. A fim de que v seja múltiplo de u, isto é, $v = \alpha u$ para algum $\alpha \in \mathbb{R}$, é necessário e suficiente que se tenha $ad - bc = 0$. A necessidade é imediata pois $v = \alpha u$ significa $c = \alpha a$ e $d = \alpha b$. Multiplicando a primeira destas igualdades por b e a segunda por a obtemos $bc = \alpha ab$ e $ad = \alpha ab$, logo $ad = bc$, ou seja, $ad - bc = 0$. Reciprocamente, se $ad = bc$ então, supondo $a \neq 0$, obtemos $d = (c/a)b$. Além disso, é claro que $c = (c/a)a$. Logo, pondo $\alpha = c/a$, vem $d = \alpha b$ e $c = \alpha a$,. isto é $v = \alpha u$. Se for $b \neq 0$, tomaremos $\alpha = d/b$ para ter $v = \alpha u$.

Exercícios

1.1. Dadas as matrizes

$$\mathbf{a} = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 2 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 & 3 & 0 \\ -2 & -3 & 1 \end{bmatrix}$$

e

$$\mathbf{c} = \begin{bmatrix} -4 & -8 & 4 \\ 12 & 13 & 1 \end{bmatrix} :$$

(a) Calcule a matriz $3\mathbf{a} - 2\mathbf{b} + \mathbf{c}$.

(b) Ache números α e β, ambos diferentes de zero, tais que $\alpha\mathbf{a}+\beta\mathbf{b}+\mathbf{c}$ tenha a primeira coluna nula.

1.2. Mostre que as operações definidas no texto fazem realmente dos conjuntos $\mathbb{R}^n$, $M(m \times n)$ e $\mathcal{F}(X; \mathbb{R})$ espaços vetoriais.

1.3. Ache o valor de t que torna a matriz abaixo igual à matriz nula:

$$\begin{bmatrix} t^2 - 1 & t^2 - t \\ t^3 - 1 & t^2 - 3t + 2 \end{bmatrix} .$$

1.4. Determine os vetores $u, v \in \mathbb{R}^4$ sabendo que as coordenadas de u são todas iguais, a última coordenada de v é igual a 3 e $u + v = (1, 2, 3, 4)$.

1.5. Dados $u = (1, 2, 3)$, $v = (3, 2, 0)$ e $w = (2, 0, 0)$, ache números α, β e γ tais que $\alpha u + \beta v + \gamma w = (1, 1, 1)$.

1.6. Dados os vetores $v_1 = (1,2,1)$, $v_2 = (2,1,2)$, $v_3 = (3,3,2)$ e $v_4 = (1,5,-1)$ em $\mathbb{R}^3$, determine os vetores $u = v_1 - 3v_2 + 2v_3 - v_4$, $v = v_1 + v_2 - v_3 - v_4$ e $w = v_3 - \frac{1}{3}v_2 - \frac{4}{3}v_1$.

1.7. Considere a seguinte afirmação: "Num espaço vetorial E existe um único vetor nulo e cada elemento de E possui um único inverso". Qual fato demonstrado nesta seção assegura que esta afirmação é verdadeira?

1.8. Use os axiomas do espaço vetorial E para provar que, se $v \in E$ e n é um número natural então $n \cdot v = v + \cdots + v$ (n parcelas).

1.9. Sejam u, v vetores não-nulos do espaço vetorial E. Prove que v é múltiplo de u se, e somente se, u é múltiplo de v. Que se pode dizer caso não suponhamos u e v ambos diferentes de zero?

1.10. Sejam $u = (x_1, \ldots, x_n)$ e $v = (y_1, \ldots, y_n)$ vetores em $\mathbb{R}^n$. Prove que um deles é múltiplo do outro se, e somente se, $x_i y_j = x_j y_i$ para quaisquer $i, j = 1, \ldots, n$.

1.11. Use as relações $2(u + v) = 2u + 2v$, $2w = w + w$ para provar que a comutatividade $u + v = v + u$ pode ser demonstrada a partir dos demais axiomas de espaço vetorial.

1.12. Em $\mathbb{R}^2$, mantenhamos a definição do produto αv de um número por um vetor mas modifiquemos, de 3 maneiras diferentes, a definição da soma $u + v$ dos vetores $u = (x, y)$ e $v = (x', y')$. Em cada tentativa, dizer quais axiomas de espaço vetorial continuam válidos e quais são violados:

(1) $u + v = (x + y', x' + y)$;

(2) $u + v = (xx', yy')$;

(3) $u + v = (3x + 3x', 5x + 5x')$.

1.13. Defina a *média* $u * v$ entre dois vetores u, v no espaço vetorial E pondo $u * v = \frac{1}{2}u + \frac{1}{2}v$. Prove que $(u * v) * w = u * (v * w)$ se, e somente se, $u = w$.

1.14. Dados os espaços vetoriais E_1, E_2, considere o conjunto $E = E_1 \times E_2$ (produto cartesiano de E_1 por E_2), cujos elementos são os pares ordenados $v = (v_1, v_2)$, com $v_1 \in E_1$ e $v_2 \in E_2$. Defina operações que tornem E um espaço vetorial. Verifique a validez de cada um dos axiomas e mostre que sua definição se estende para o caso de

n espaços vetoriais $E_1, \ldots, E_n$, ou mesmo de uma seqüência infinita $E_1, E_2, \ldots, E_n, \ldots$.

1.15. Sejam X um conjunto qualquer e E um espaço vetorial. Mostre que, com as definições naturais, o conjunto $\mathcal{F}(X; E)$ das funções $f \colon X \to E$ se torna um espaço vetorial. Identifique os casos particulares em que $X = \{1, \ldots, n\}$, $X = \mathbb{N}$, $X = A \times B$, onde $A = \{1, \ldots, m\}$ e $B = \{1, \ldots, n\}$.

1.16. Dados os vetores $u = (1, 2, 3)$, $v = (3, 2, 1)$ e $w = (-3, 2, 7)$ em $\mathbb{R}^3$, obtenha números α, β tais que $w = \alpha u + \beta v$. Quantas soluções admite este problema?

1.17. Sejam $u = (1, 1)$, $v = (1, 2)$ e $w = (2, 1)$. Ache números a, b, c, a', b', c', todos não-nulos, tais que $au + bv + cw = a'u + b'v + c'w$, com $a' \neq a$, $b' \neq b$, $c' \neq c$.

1.18. Sejam E um espaço vetorial e $u, v \in E$. O *segmento de reta* de extremidades u, v é, por definição, o conjunto

$$[u, v] = \{(1 - t)u + tv; \;\; 0 \leq t \leq 1\}.$$

Um conjunto $X \subset E$ chama-se *convexo* quando $u, v \in X \Rightarrow [u, v] \subset X$. (Ou seja: o segmento de reta que liga dois pontos quaisquer de X está contido em X.) Prove:

(a) A interseção $X_1 \cap \ldots \cap X_m$ de conjuntos convexos $X_1, \ldots, X_m \subset E$ é um conjunto convexo.

(b) Dados $a, b, c \in \mathbb{R}$, o conjunto $X = \{(x, y) \in \mathbb{R}^2; ax + by \leq c\}$ é convexo em $\mathbb{R}^2$.

(c) O conjunto $Y = \{(x, y, z) \in \mathbb{R}^3; \; a \leq x \leq b, \; c < y < d\}$ é convexo em $\mathbb{R}^3$.

(d) Seja $X \subset E$ convexo. Se r, s, t são números reais ≥ 0 tais que $r + s + t = 1$ então $u, v, w \in X \Rightarrow ru + sv + tw \in X$.

(e) Generalizando o resultado acima, a expressão $t_1 v_1 + \cdots + t_k v_k$, onde $t_1, \ldots, t_k$ são ≥ 0 e $t_1 + \cdots + t_k = 1$ chama-se uma *combinação convexa* dos vetores $v_1, \ldots, v_k$. Se o conjunto $X \subset E$ é convexo, prove que toda combinação convexa de vetores $v_1, \ldots, v_k \in X$ ainda pertence a X.

1.19. Prove que o disco $D = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}$ é um conjunto convexo.

1.20. Um subconjunto C do espaço vetorial E chama-se um *cone* quando, para todo $v \in C$ e todo $t > 0$, tem-se $tv \in C$. Prove:

(a) O conjunto dos vetores $v \in \mathbb{R}^n$ que têm exatamente k coordenadas positivas $(0 \leq k \leq n)$ é um cone.

(b) O conjunto das funções $f\colon X \to \mathbb{R}$ que assumem valores negativos em todos os pontos de um subconjunto fixado $Y \subset X$ é um cone em $\mathcal{F}(X; \mathbb{R})$.

(c) Um cone $C \subset E$ é um conjunto convexo se, e somente se, $u, v \in C \Rightarrow u + v \in C$.

(d) A interseção e a reunião de uma família qualquer de cones são ainda cones.

1.21. Dado um subconjunto X no espaço vetorial E, seja $C(X)$ o conjunto das combinações convexas $t_1 v_1 + \cdots + t_k v_k$ $(t_i \geq 0, \Sigma t_i = 1)$ dos elementos de X. Prove que $C(X)$ é um conjunto convexo, que $X \subset C(X)$ e que se C' é qualquer subconjunto convexo de E contendo X então $C' \supset C(X)$. (Por este motivo, diz-se que $C(X)$ é o menor subconjunto convexo de E que contém X. $C(X)$ chama-se a *envoltória convexa* do conjunto X.)

2

Subespaços

Um subespaço vetorial do espaço vetorial E é um subconjunto $F \subset E$ que, relativamente às operações de E, é ainda um espaço vetorial. Os subespaços vetoriais constituem uma rica fonte de exemplos de espaços vetoriais, como se verá nas seções seguintes.

Seja E um espaço vetorial. Um *subespaço vetorial* (ou simplesmente um *subespaço*) de E é um subconjunto $F \subset E$ com as seguintes propriedades:

1. $0 \in F$;
2. Se $u, v \in F$ então $u + v \in F$;
3. Se $v \in F$ então, para todo $\alpha \in \mathbb{R}$, $\alpha v \in F$.

Segue-se que se u e v pertencem ao subespaço F e α, β são números reais quaisquer então $\alpha u + \beta v \in F$. Mais geralmente, dados $v_1, \ldots, v_m \in F$ e $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ tem-se $v = \alpha_1 v_1 + \cdots + \alpha_m v_m \in F$.

O conjunto $\{0\}$, com o único elemento 0, e o espaço inteiro E são exemplos triviais de subespaços de E. Todo subespaço é, em si mesmo, um espaço vetorial.

Exemplo 2.1. Seja $v \in E$ um vetor não-nulo. O conjunto $F = \{\alpha v; \alpha \in \mathbb{R}\}$ de todos os múltiplos de v é um subespaço vetorial de E, chamado a *reta que passa pela origem e contém* v.

Exemplo 2.2. Seja $E = \mathcal{F}(\mathbb{R}; \mathbb{R})$ o espaço vetorial das funções reais de uma variável real $f \colon \mathbb{R} \to \mathbb{R}$. Para cada $k \in \mathbb{N}$, o conjunto $C^k(\mathbb{R})$

das funções k vezes continuamente deriváveis é um subespaço vetorial de E. Também são subespaços de E o conjunto $C^o(\mathbb{R})$ das funções contínuas, o conjunto $C^\infty(\mathbb{R})$ das funções infinitamente deriváveis, o conjunto $\mathcal{P} = \mathcal{P}(\mathbb{R})$ dos polinômios $p(x) = a_o + a_1 x + \cdots + a_n x^n$ e o conjunto $\mathcal{P}_n$ dos polinômios de grau $\leq n$. Para $n, k \in \mathcal{N}$ quaisquer, tem-se:

$$C^o(\mathbb{R}) \supset C^k(\mathbb{R}) \supset C^{k+1}(\mathbb{R}) \supset C^\infty(\mathbb{R}) \supset \mathcal{P} \supset \mathcal{P}_n.$$

Observe que o conjunto dos polinômios de grau n não é um subespaço vetorial de E pois a soma de dois polinômios de grau n pode ter grau $< n$.

Exemplo 2.3. Sejam $a_1, \ldots, a_n$ números reais. O conjunto H de todos os vetores $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$ tais que

$$a_1 x_1 + \cdots + a_n x_n = 0$$

é um subespaço vetorial de $\mathbb{R}^n$. No caso desinteressante em que $a_1 = \cdots = a_n = 0$, o subespaço H é todo o $\mathbb{R}^n$. Se, ao contrário, pelo menos um dos a_i é $\neq 0$, H chama-se um *hiperplano* de $\mathbb{R}^n$ que passa pela origem.

Exemplo 2.4. Sejam E um espaço vetorial e L um conjunto de índices. Se, para cada $\lambda \in L$, F_λ é um subespaço vetorial de E então a interseção

$$F = \bigcap_{\lambda \in L} F_\lambda$$

é ainda um subespaço vetorial de E. Segue-se então do Exemplo 2.3 que o conjunto dos vetores $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$ cujas coordenadas satisfazem as m condições abaixo

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

é um subespaço vetorial de $\mathbb{R}^n$, o qual é a interseção $F = F_1 \cap \ldots \cap F_m$ dos hiperplanos F_i definidos, segundo o Exemplo 2.3, por cada uma das equações acima.

Seja X um subconjunto do espaço vetorial E. O *subespaço vetorial de* E *gerado por* X é, por definição, o conjunto de todas as combinações lineares

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m$$

de vetores $v_1, \ldots, v_m \in X$.

É fácil ver que o conjunto de todas as combinações lineares que se podem formar com vetores retirados do conjunto X é, de fato, um subespaço vetorial, que indicaremos pelo símbolo $S(X)$.

O subespaço $S(X)$, gerado pelo subconjunto $X \subset E$, contém o conjunto X e, além disso, é o menor subespaço de E que contém X. Noutras palavras, se F é um subespaço vetorial de E e $X \subset F$ então $S(X) \subset F$. Evidentemente, se X já é um subespaço vetorial, então $S(X) = X$. Quando o subespaço $S(X)$ coincide com E, diz-se que X é um *conjunto de geradores de* E.

Explicitamente: um conjunto X é um conjunto de geradores do espaço vetorial E quando todo vetor $w \in E$ pode exprimir-se como combinação linear

$$w = \alpha_1 v_1 + \cdots + \alpha_m v_m$$

de vetores $v_1, \ldots, v_m$ pertencentes a X.

Exemplo 2.5. Se $v \in E$ é um vetor não-nulo, o subespaço gerado por v é a reta que passa pela origem e contém v.

Exemplo 2.6. Sejam $u = (a, b)$ e $v = (c, d)$ vetores de $\mathbb{R}^2$ tais que nenhum deles é múltiplo do outro. Então $u \neq 0$, $v \neq 0$ e, pelo Exemplo 1.5, $ad - bc \neq 0$. Afirmamos que $X = \{u, v\}$ é um conjunto de geradores de $\mathbb{R}^2$, ou seja, que qualquer vetor $w = (r, s) \in \mathbb{R}^2$ pode exprimir-se como uma combinação linear $w = xu + yv$. De fato esta igualdade vetorial em $\mathbb{R}^2$ equivale às duas igualdades numéricas

$$\begin{aligned} ax + cy &= r \\ bx + dy &= s. \end{aligned}$$

Como $ad - bc \neq 0$, o sistema de equações acima possui uma solução (x, y), logo existem $x, y \in \mathbb{R}$ tais que $xu + yv = w$. Esta mesma conclusão pode também ser obtida geometricamente, conforme mostra a Figura 2.1. A partir da ponta de w, traçam-se paralelas às retas que contêm u e v, determinando assim os múltiplos xu, yv, que somados dão w.

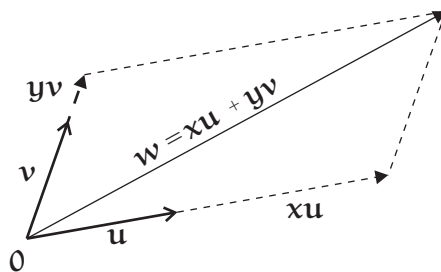



Figura 2.1.

Exemplo 2.7. Os chamados *vetores canônicos*

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0), \\ e_2 &= (0, 1, 0, \ldots, 0), \\ &\vdots \\ e_n &= (0, 0, \ldots, 0, 1) \end{aligned}$$

constituem um conjunto de geradores do espaço $\mathbb{R}^n$. Com efeito, dado $v = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$, tem-se $v = \alpha_1 e_1 + \cdots + \alpha_n e_n$. Analogamente, os monômios $1, x, \ldots, x^n, \ldots$ (em número infinito) formam um conjunto de geradores do espaço $\mathcal{P}$ dos polinômios reais. Por sua vez, os $n+1$ primeiros deles, a saber, $1, x, \ldots, x^n$ constituem um conjunto de geradores de $\mathcal{P}_n$, espaço vetorial dos polinômios de grau $\leq n$.

Resulta do Exemplo 2.6 que os únicos subespaços vetoriais de $\mathbb{R}^2$ são $\{0\}$, as retas que passam pela origem e o próprio $\mathbb{R}^2$. Com efeito, seja $F \subset \mathbb{R}^2$ um subespaço vetorial. Se F contém apenas o vetor nulo, então $F = \{0\}$. Se F contém algum vetor $u \neq 0$ então há duas possibilidades: ou todos os demais vetores de F são múltiplos de u, e neste caso F é a reta que passa pela origem e contém u, ou então F contém, além de u, um outro vetor v que não é múltiplo de u. Neste caso, F contém todas as combinações lineares $xu + yv$, logo $F = \mathbb{R}^2$, pelo Exemplo 2.6.

Exemplo 2.8. O sistema linear de m equações a n incógnitas

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

possui uma solução $(x_1, \ldots, x_n)$ se, e somente se, o vetor $b = (b_1, \ldots, b_m)$ é combinação linear dos vetores-coluna

$$\begin{aligned} v_1 &= (a_{11}, a_{21}, \ldots, a_{m1}), \\ &\vdots \\ v_n &= (a_{1n}, a_{2n}, \ldots, a_{mn}) \end{aligned}$$

da matriz $\mathbf{a} = [a_{ij}]$. Com efeito, estas equações significam que

$$b = x_1v_1 + x_2v_2 + \cdots + x_nv_n.$$

Em particular, se os vetores-coluna $v_1, \ldots, v_n$ gerarem $\mathbb{R}^m$, o sistema possui solução, seja qual for o segundo membro b.

Sejam F_1 e F_2 subespaços vetoriais de E. O subespaço vetorial de E gerado pela reunião $F_1 \cup F_2$ é, como se vê facilmente, o conjunto de todas as somas $v_1 + v_2$, onde $v_1 \in F_1$ e $v_2 \in F_2$. Ele é representado pelo símbolo $F_1 + F_2$.

Mais geralmente, dados os subconjuntos $X, Y \subset E$, indica-se com $X + Y$ o conjunto cujos elementos são as somas $u + v$, onde $u \in X$ e $v \in Y$. Quando $X = \{u\}$ reduz-se a um único elemento u, escreve-se $u + Y$ em vez de $\{u\} + Y$. Diz-se então que $u + Y$ resulta de Y pela translação de u.

Quando os subespaços $F_1, F_2 \subset E$ têm em comum apenas o elemento $\{0\}$, escreve-se $F_1 \oplus F_2$ em vez de $F_1 + F_2$ e diz-se que $F = F_1 \oplus F_2$ é a *soma direta* de F_1 e F_2.

Teorema 2.1. *Sejam F, F_1, F_2 subespaços vetoriais de E, com $F_1 \subset F$ e $F_2 \subset F$. As seguintes afirmações são equivalentes*:

(1) $F = F_1 \oplus F_2$;

(2) *Todo elemento* $w \in F$ *se escreve, de modo único, como soma* $w = v_1 + v_2$, *onde* $v_1 \in F_1$ *e* $v_2 \in F_2$.

Demonstração: Provemos que (1) $\Rightarrow$ (2). Para isto, suponhamos que $F_1 \cap F_2 = \{0\}$ e que se tenha $u_1 + u_2 = v_1 + v_2$, com $u_1, v_1 \in F_1$ e $u_2, v_2 \in F_2$. Então $u_1 - v_1 = v_2 - u_2$. Como $u_1 - v_1 \in F_1$ e $v_2 - u_2 \in F_2$, segue-se que $u_1 - v_1$ e $v_2 - u_2$ pertencem ambos a F_1 e a F_2. Mas $F_1 \cap F_2 = \{0\}$. Logo $u_1 - v_1 = v_2 - u_2 = 0$, ou seja, $u_1 = v_1$ e $u_2 = v_2$. Para provar que (2) $\Rightarrow$ (1), seja $v \in F_1 \cap F_2$. Então $0 + v = v + 0$ com $0, v \in F_1$ e $v, 0 \in F_2$. Pela hipótse (2), isto implica $0 = v$, portanto $F_1 \cap F_2 = \{0\}$. □

Exemplo 2.9. Em $\mathbb{R}^4$, sejam F_1 o subespaço gerado pelos vetores $e_1 = (1, 0, 0, 0)$, $e_3 = (0, 0, 1, 0)$ e F_2 o subespaço gerado pelos vetores $e_2 = (0, 1, 0, 0)$, $e_4 = (0, 0, 0, 1)$. Então F_1 é o conjunto dos vetores da forma $(\alpha_1, 0, \alpha_3, 0)$ enquanto os vetores de F_2 têm a forma $(0, \alpha_2, 0, \alpha_4)$. É claro que $\mathbb{R}^4 = F_1 \oplus F_2$.

A noção de subespaço vetorial abrange as retas, planos e seus análogos multidimensionais apenas nos casos em que esses conjuntos contêm a origem. Para incluir retas, planos, etc. que não passam pela origem, tem-se a noção de variedade afim, que discutiremos agora.

Seja E um espaço vetorial. Se $x, y \in E$ e $x \neq y$, a *reta* que une os pontos x, y é, por definição o conjunto

$$r = \{(1 - t)x + ty; t \in \mathbb{R}\}.$$

Pondo $v = y - x$, podemos ver que $r = \{x + tv; t \in \mathbb{R}\}$.

Um subconjunto $V \subset E$ chama-se uma *variedade afim* quando a reta que une dois pontos quaisquer de V está contida em V. Assim, $V \subset E$ é uma variedade afim se, e somente se, cumpre a seguinte condição:

$$x, y \in V, t \in \mathbb{R} \Rightarrow (1 - t)x + ty \in V.$$

Exemplo 2.10. Um exemplo óbvio de variedade afim é um subespaço vetorial. Ao contrário dos subespaços vetoriais, que nunca são vazios pois devem conter o zero, a definição acima é formulada de tal modo que o conjunto vazio a cumpre, logo $\varnothing$ é uma variedade afim.

Se $V_1, \ldots, V_m \subset E$ são variedades afins então a interseção $V = V_1 \cap \ldots \cap V_m$ é ainda uma variedade afim. Todo ponto $p \in E$ é uma variedade afim.

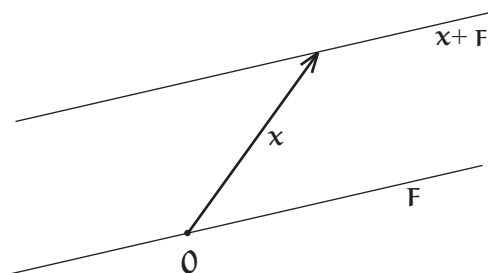
Exemplo 2.11. Sejam $a_1, \ldots, a_n, b$ números reais. O conjunto H dos pontos $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ tais que

$$a_1x_1 + \cdots + a_nx_n = b$$

é uma variedade afim (que não contém a origem quando $b \neq 0$). Se os números a_i não são todos nulos, H chama-se um hiperplano. Se $a_1 = \cdots = a_n = 0$, tem-se $H = \varnothing$ quando $b \neq 0$ e $H = \mathbb{R}^n$ quando $b = 0$. Mais geralmente, o conjunto das soluções de um sistema linear de m equações com n incógnitas (vide Exemplo 2.8) é uma variedade afim (eventualmente vazia), interseção das m variedades afins definidas pelas equações do sistema.

O teorema a seguir mostra que toda variedade afim não-vazia V pode ser obtida transladando-se um subespaço vetorial F. Diz-se que F é o subespaço vetorial *paralelo* a V.

Teorema 2.2. *Seja V uma variedade afim não-vazia no espaço vetorial E. Existe um único subespaço vetorial $F \subset E$ tal que, para todo $x \in V$ tem-se*

$$V = x + F = \{x + v; v \in F\}.$$

Demonstração: Dado $x \in V$, seja F o conjunto de todos os vetores $v = y - x$, onde $y \in V$. Mostremos que F é um subespaço vetorial. É claro que $0 \in F$. Além disso, se $\alpha \in \mathbb{R}$ e $v \in F$ então $v = y - x$, com $y \in V$, logo

$$\alpha v = \alpha(y - x) = [(1 - \alpha)x + \alpha y] - x = z - x,$$

com $z = (1-\alpha)x + \alpha y \in V$. Portanto $\alpha v \in F$. Finalmente, se $v = y - x$ e $v' = y' - x$ pertencem a F então

$$z = \frac{1}{2}y + \frac{1}{2}y' \in V,$$

portanto $z - x \in F$. Segue-se daí que a soma

$$v + v' = y + y' - 2x = 2(z - x)$$

pertence a F.

Em seguida, mostremos que $V = x + F$. Com efeito, $y \in V \Rightarrow y = x + (y - x)$ com $y - x \in F$, logo $y \in x + F$. Assim, $V \subset x + F$. Por outro lado, um elemento qualquer de $x + F$ tem a forma $x + (y - x)$, com $y \in V$, logo é igual a y e daí $x + F \subset V$.

Finalmente, se F e F' são subespaços vetoriais de E, tais que $x + F = x + F'$ para algum $x \in E$, provemos que se tem $F = F'$. Com efeito, $v \in F \Rightarrow x + v \in x + F \Rightarrow x + v \in x + F' \Rightarrow x + v = x + v' (v' \in F') \Rightarrow v = v' \Rightarrow v \in F'$. Portanto $F \subset F'$. Da mesma forma, vê-se que $F' \subset F$, o que conclui a demonstração. □

Exemplo 2.12. Vimos no exemplo 2.8 que o conjunto V das soluções de um sistema linear de m equações com n incógnitas é uma variedade afim. Supondo $V \neq \varnothing$, tomemos $x_0 \in V$ e chamemos de F o subespaço vetorial de $\mathbb{R}^n$ formado pelas soluções do sistema homogêneo correspondente (descrito no Exemplo 2.4; veja também a página 27). Tem-se $V = x_0 + F$. Diz-se então que "todas as soluções do sistema se obtêm somando uma solução particular com a solução geral do sistema homogêneo associado".

Exercícios

2.1. Seja $\mathbb{R}^{(\infty)}$ o subconjunto de $\mathbb{R}^\infty$ formado pelas seqüências $v = (x_1, x_2, \ldots, x_n, \ldots)$ que têm apenas um número finito de termos x_n diferentes de zero. Mostre que $\mathbb{R}^{(\infty)}$ é um subespaço vetorial de $\mathbb{R}^\infty$ e que as seqüências que têm um único termo não-nulo constituem um conjunto de geradores para $\mathbb{R}^{(\infty)}$.

2.2. Use o índice deste livro para localizar a definição de matriz triangular. Mostre que o conjunto F_1 das matrizes triangulares inferiores e o conjunto F_2 das matrizes triangulares superiores são

subespaços vetoriais de $M(n \times n)$, que $M(n \times n) = F_1 + F_2$ e que não se tem $M(n \times n) = F_1 \oplus F_2$.

2.3. Seja $E = \mathcal{F}(\mathbb{R};\mathbb{R})$. Para $X \subset \mathbb{R}$ qualquer, ponhamos $N(X) = \{\varphi \in E;\ \varphi(x) = 0$ para todo $x \in X\}$. Prove:

(a) Para todo $X \subset \mathbb{R}$, $N(X)$ é um subespaço vetorial de E.

(b) $X \subset Y \Rightarrow N(Y) \subset N(X)$

(c) $N(X \cup Y) = N(X) \cap N(Y)$

(d) $N(X) = \{0\} \Leftrightarrow X = \mathbb{R}$

(e) $N(X \cap Y) = N(X) + N(Y)$

(f) $N(X) \oplus N(Y) = E \Leftrightarrow Y = \mathbb{R} - X$.

2.4. No espaço vetorial $E = \mathcal{F}(\mathbb{R};\mathbb{R})$ sejam:

F_1 = conjunto das funções $f\colon \mathbb{R} \to \mathbb{R}$ que se anulam em todos os pontos do intervalo $[0,1]$;
F_2 = conjunto das funções $g\colon \mathbb{R} \to \mathbb{R}$ que se anulam em todos os pontos do intervalo $[2,3]$.

Mostre que F_1 e F_2 são subespaços vetoriais de E, que $E = F_1 + F_2$ e que não se tem $E = F_1 \oplus F_2$.

2.5. Considere os subespaços $F_1, F_2 \subset \mathbb{R}^3$ assim definidos: F_1 é o conjunto de todos os vetores $v = (x,x,x)$ que têm as três coordenadas iguais e F_2 é o conjunto de todos os vetores $w = (x,y,0)$ que têm a última coordenada igual a zero. Mostre que $\mathbb{R}^3 = F_1 \oplus F_2$.

2.6. Dados $u = (1,2)$ e $v = (-1,2)$, sejam F_1 e F_2 respectivamente as retas que passam pela origem em $\mathbb{R}^2$ e contêm u e v. Mostre que $\mathbb{R}^2 = F_1 \oplus F_2$.

2.7. Sejam $F_1 = S(u_1,v_1)$ e $F_2 = S(u_2,v_2)$ os subespaços de $\mathbb{R}^3$ gerados pelos vetores $u_1 = (0,1,-2)$, $v_1 = (1,1,1)$, $u_2 = (-1,0,3)$ e $v_2 = (2,-1,0)$. Ache números a_1, b_1, c_1 e a_2, b_2, c_2 tais que se tenha:

$$F_1 = \{(x,y,z) \in \mathbb{R}^3; a_1x + b_1y + c_1z = 0\}$$
$$F_2 = \{(x,y,z) \in \mathbb{R}^3; a_2x + b_2y + c_2z = 0\}.$$

2.8. No exercício anterior, mostre que $u_2 \notin F_1$ e que $F_1 + F_2 = \mathbb{R}^3$. Exiba um vetor não nulo $w \in F_1 \cap F_2$ e conclua que não se tem $\mathbb{R}^3 = F_1 \oplus F_2$.

2.9. Prove que $S(X)$ é a interseção de todos os subespaços vetoriais que contêm o conjunto $X \subset E$.

2.10. Exiba três vetores $u, v, w \in \mathbb{R}^3$ com as seguintes propriedades: nenhum deles é múltiplo do outro, nenhuma das coordenadas é igual a zero e $\mathbb{R}^3$ não é gerado por eles.

2.11. Seja F o subespaço de $\mathbb{R}^3$ gerado pelos vetores $u = (1, 1, 1)$ e $v = (1, -1, -1)$. Ache números a, b, c com a seguinte propriedade: um vetor $w = (x, y, z)$ pertence a F se, e somente se, $ax + by + cz = 0$.

2.12. Exprima o vetor $(1, -3, 10)$ como combinação linear dos vetores $u = (1, 0, 0)$, $v = (1, 1, 0)$ e $w = (2, -3, 5)$.

2.13. Mostre que a matriz $\mathbf{d} = \begin{bmatrix} 4 & -4 \\ -6 & 16 \end{bmatrix}$ pode ser escrita como combinação linear das matrizes

$$\mathbf{a} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 & 2 \\ 3 & -4 \end{bmatrix} \quad \text{e} \quad \mathbf{c} = \begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix}.$$

2.14. Assinale V(erdadeiro) ou F(also):

( ) O vetor $w = (1, -1, 2)$ pertence ao subespaço gerado por $u = (1, 2, 3)$ e $v = (3, 2, 1)$.

( ) Qualquer vetor em $\mathbb{R}^3$ pode ser expresso como combinação linear dos vetores $u = (-5, 3, 2)$ e $v = (3, -1, 3)$.

( ) Se $X \subset Y$ então $S(X) \subset S(Y)$.

( ) Se $S(X) \subset S(Y)$ então $X \subset Y$.

( ) Se uma variedade afim $V \subset E$ contém o vetor zero então V é um subespaço vetorial de E.

2.15. Quais dos seguintes subconjuntos são subespaços vetoriais?

(a) O conjunto $X \subset \mathbb{R}^3$ formado pelos vetores $v = (x, y, z)$ tais que $z = 3x$ e $x = 2y$.

(b) O conjunto $Y \subset \mathbb{R}^3$ formado pelos vetores $v = (x, y, z)$ tais que $xy = 0$.

(c) O conjunto Z das matrizes 2×3 nas quais alguma coluna é formada por elementos iguais.

(d) O conjunto $F \subset \mathcal{F}(\mathbb{R};\mathbb{R})$ formado pelas funções $f\colon \mathbb{R} \to \mathbb{R}$ tais que $f(x+1) = f(x)$ para todo $x \in \mathbb{R}$.

(e) O conjunto $L \subset \mathbb{R}^n$ dos vetores $v = (x, 2x, \ldots, nx)$, onde $x \in \mathbb{R}$ é arbitrário.

(f) O conjunto dos vetores $v \in \mathbb{R}^5$ que têm duas ou mais coordenadas nulas.

(g) O conjunto dos vetores de $\mathbb{R}^3$ que têm pelo menos uma coordenada ≥ 0.

2.16. Exprima, em termos das operações num espaço vetorial E, uma condição para que $u, v, w \in E$ sejam colineares (isto é, pertençam a uma mesma reta, que pode conter ou não o vetor zero).

2.17. Obtenha números a, b, c, d tais que a variedade afim (plano) de $\mathbb{R}^3$ definida pela equação $ax + by + cz = d$ contenha os pontos $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ e $e_3 = (0, 0, 1)$.

2.18. Prove que, na definição de subespaço vetorial, a condição "$0 \in F$" pode ser substituída por "$F \neq \varnothing$".

2.19. Quais dos seguintes conjuntos são subespaços vetoriais?
(a) O conjunto dos vetores de $\mathbb{R}^n$ cujas coordenadas formam uma progressão aritmética.
(b) Os vetores de $\mathbb{R}^n$ cujas coordenadas formam uma progressão geométrica.
(c) Os vetores de $\mathbb{R}^n$ cujas coordenadas formam uma progressão aritmética de razão fixada.
(d) Os vetores de $\mathbb{R}^n$ cujas coordenadas formam uma progressão geométrica de razão fixada.
(e) Os vetores de $\mathbb{R}^n$ cujas primeiras k coordenadas são iguais.
(f) Os vetores de $\mathbb{R}^n$ que têm k coordenadas iguais.
(g) As seqüências $(x_n) \in \mathbb{R}^\infty$ tais que $x_{n+2} - 3x_n = x_{n+1}$ para todo n.
(h) Os vetores $(x, y) \in \mathbb{R}^2$ tais que $x^2 + 3x = y^2 + 3y$.
(i) As funções $f \in C^\infty(\mathbb{R})$ tais que $f'' - 2f' + f = 0$.

2.20. Sejam v_1, v_2, v_3 os vetores-linha e w_1, w_2, w_3 os vetores-coluna da matriz

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Verifique as relações $v_3 = 2v_2 - v_1$, $w_3 = 2w_2 - w_1$. Exprima w_1 e w_2 como combinações lineares de v_1 e v_2, e vice-versa. Conclua que os vetores-linha e os vetores-coluna da matriz dada geram o mesmo subespaço de $\mathbb{R}^3$.

2.21. Dê exemplo de uma matriz 3×3 cujos vetores-linha geram um subespaço de $\mathbb{R}^3$ diferente daquele gerado pelos vetores-coluna.

2.22. Prove que a reunião de dois subespaços vetoriais de E é um subespaço vetorial se, e somente se, um deles estiver contido no outro.

2.23. A partir da definição, prove que, dados os números $a_1, ..., a_n, c$, o conjunto V dos vetores $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ tais que $a_1x_1 + \cdots + a_nx_n = c$ é um subespaço vetorial de $\mathbb{R}^n$ se, e somente se, $c = 0$. Prove a afirmação feita no texto de que V é uma variedade afim.

2.24. Seja F um subespaço vetorial de E. Assinale V(erdadeiro) ou F(also):

( ) Se $u \notin F$ e $v \notin F$ então $u + v \notin F$;

( ) Se $u \notin F$ e $\alpha \neq 0$ então $\alpha u \notin F$.

2.25. Diz-se que um subconjunto X de um espaço vetorial E é *simétrico* quando $v \in X \Rightarrow -v \in X$. Prove que um cone convexo simétrico e não-vazio é um subespaço vetorial de E.

2.26. Dê exemplo de um cone convexo que não seja simétrico e um cone simétrico que não seja convexo.

2.27. Uma matriz quadrada $\mathbf{a} = [a_{ij}]$ chama-se *simétrica* (respect. *anti-simétrica*) quando $a_{ij} = a_{ji}$ (respect. $a_{ij} = -a_{ji}$) para todo i e todo j. Prove que o conjunto S das matrizes simétricas e o conjunto A das matrizes anti-simétricas $n \times n$ são subespaços vetoriais de $M(n \times n)$ e que se tem $M(n \times n) = S \oplus A$.

2.28. Seja $E = \mathcal{F}(\mathbb{R}; \mathbb{R})$. Fixada $g \colon \mathbb{R} \to \mathbb{R}$, mostre que o conjunto F de todas as funções $f \colon \mathbb{R} \to \mathbb{R}$ tais que $f(g(x)) = f(x)$ é um subespaço

vetorial de E. Para qual função g tem-se $F =$ conjunto das funções periódicas de período a? E se fosse $g(f(x)) = f(x)$? Ou $f(g(x)) = g(x)$?

2.29. Prove que o subespaço vetorial gerado por um cone convexo $C \subset E$ é o conjunto das diferenças $u - v$, onde $u, v \in C$. Conclua que o conjunto das funções $f \colon X \to \mathbb{R}$ que só assumem valores positivos é um conjunto de geradores de $\mathcal{F}(X; \mathbb{R})$.

2.30. Diz-se que uma função $f \colon X \to \mathbb{R}$ é *limitada* quando existe $k > 0$ (dependendo de f) tal que $|f(x)| \leq k$ para todo $x \in X$. Prove que o conjunto das funções limitadas é um subespaço vetorial de $\mathcal{F}(X; \mathbb{R})$, o qual é gerado pelas funções limitadas positivas.

2.31. Um subespaço vetorial de $\mathbb{R}^3$ gerado por dois vetores não-colineares u, v chama-se um *plano*. Use um argumento geométrico para provar que se o vetor $w \in \mathbb{R}^3$ não pertence ao plano gerado por u e v então u, v e w geram $\mathbb{R}^3$.

2.32. Mostre que o vetor $b = (1, 2, 2)$ não é combinação linear dos vetores $v_1 = (1, 1, 2)$ e $v_2 = (1, 2, 1)$. A partir daí, formule um sistema linear de 3 equações com 2 incógnitas, que não possui solução e que tem o vetor b como segundo membro.

2.33. Sejam $F_1, \ldots, F_k \subset E$ subespaços vetoriais. Prove:

(1) O subespaço gerado pela união $F_1 \cup \ldots \cup F_k$ é o conjunto $F_1 + \cdots + F_k$ das somas $x_1 + \cdots + x_k$, onde $x_1 \in F_1, \ldots, x_k \in F_k$.

(2) As seguintes afirmações são equivalentes:

(a) Cada $x \in F_1 + \cdots + F_k$ se escreve de modo único como soma $x = x_1 + \cdots + x_k$.

(b) Para cada $j = 1, \ldots, k$ tem-se $F_j \cap (F_1 + \cdots + F_{j-1} + F_{j+1} + \cdots + F_k) = \{0\}$.

Quando uma das condições (a) ou (b) vale, escreve-se $F_1 \oplus \cdots \oplus F_k$ em vez de $F_1 + \cdots + F_k$ e diz-se que este subespaço é a *soma direta* de $F_1, \ldots, F_k$.

2.34. Seja $E = F_1 \oplus F_2 = G_1 \oplus G_2$. Se $F_1 \subset G_1$ e $F_2 \subset G_2$, prove que $F_1 = G_1$ e $F_2 = G_2$.

2.35. Sejam E, F espaços vetoriais. Uma função $f\colon E \to F$ chama-se *par* (respect. *ímpar*) quando $f(-v) = f(v)$ (respect. $f(-v) = -f(v)$) para todo $v \in E$. Prove:

(a) O conjunto A das funções pares e o conjunto B das funções ímpares são subespaços vetoriais de $\mathcal{F}(E;F)$ (vide Exerc. 1.15) e vale $\mathcal{F}(E;F) = A \oplus B$.

(b) Além dos conjuntos A, dos polinômios pares, e B, dos polinômios ímpares, considere também o conjunto A' dos polinômios da forma $p(x) = \Sigma a_i x^{2i}$ que só contêm expoentes pares e o conjunto B' dos polinômios da forma $q(x) = \Sigma a_i x^{2i+1}$, que só contêm expoentes ímpares. Prove que A' e B' são subespaços vetoriais do espaço $\mathcal{P}$ de todos os polinômios, que $A' \subset A$, $B' \subset B$ e $\mathcal{P} = A' \oplus B'$. Conclua que $A = A'$ e $B = B'$.

2.36. Para todo $n \in \mathbb{N}$ seja Q_n o conjunto dos polinômios (de graus arbitrários) que são divisíveis por x^n. Prove que Q_n é um subespaço vetorial de $\mathcal{P}$. Ache um subespaço $F \subset \mathcal{P}$ tal que $\mathcal{P} = F \oplus Q_n$.

2.37. Dado $X \subset E$, seja Y o conjunto obtido de X substituindo um dos seus elementos v por $v + \alpha u$, onde $u \in X$ e $\alpha \in \mathbb{R}$. Prove que X e Y geram o mesmo subespaço vetorial de E. Conclua daí que os conjuntos $\{v_1, \ldots, v_k\} \subset E$ e $\{v_1, v_2 - v_1, \ldots, v_k - v_1\} \subset E$ geram o mesmo subespaço vetorial de E.

2.38. Prove que a reunião de três subespaços vetoriais só pode ser um subespaço vetorial quando um deles contém os outros dois.

2.39. Sejam F_1, F_2 subespaços vetoriais de E. Se existir algum $a \in E$ tal que $a + F_1 \subset F_2$, prove que $F_1 \subset F_2$.

2.40. Seja $V \subset E$ uma variedade afim. Dados $v_1, \ldots, v_m \in V$ e $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ com $\alpha_1 + \cdots + \alpha_m = 1$, prove que $\alpha_1 v_1 + \cdots + \alpha_m v_m \in V$.

2.41. Para todo subespaço vetorial $F \subset \mathbb{R}^n$, prove que existe um subespaço $G \subset \mathbb{R}^n$ tal que $\mathbb{R}^n = F \oplus G$.

2.42. Verdadeiro ou falso? Para quaisquer subconjuntos $X, Y \subset E$ tem-se

$$S(X \cup Y) = S(X) + S(Y),$$
$$S(X \cap Y) = S(X) \cap S(Y).$$

(A última das igualdades acima sugere uma pergunta: qual seria o subespaço vetorial gerado pelo conjunto vazio? A convenção mais conveniente é $S(\emptyset) = \{0\}$.)

2.43. Dado o subconjunto não-vazio X do espaço vetorial E, a *variedade afim gerada por* X é, por definição, o conjunto $V(X)$ de todas as combinações lineares $\alpha_1 v_1 + \cdots + \alpha_n v_n$, com $v_1, \ldots, v_n \in X$ e $\alpha_1 + \cdots + \alpha_n = 1$. Prove que

(a) $V(X)$ é uma variedade afim;

(b) Fixado qualquer $v_0 \in X$, tem-se $V(X) = v_0 + F$, onde F é o subespaço vetorial de E gerado pelos vetores $v - v_0$, onde $v \in X$.

3

Bases

Os espaços vetoriais de dimensão finita, objetos centrais do nosso estudo, possuem uma estrutura algébrica extremamente simples, evidenciada pelas idéias de base e dimensão, que apresentaremos agora. Uma vez fixada uma base num espaço vetorial de dimensão n, seus elementos são meramente combinações lineares dos n vetores básicos, com coeficientes univocamente determinados. Nesta seção, esses fatos serão estabelecidos e analisados em detalhe.

Seja E um espaço vetorial. Diz-se que um conjunto $X \subset E$ é *linearmente independente* (abreviadamente, L.I.) quando nenhum vetor $v \in X$ é combinação linear de outros elementos de X. Para evitar ambigüidade, no caso em que $X = \{v\}$ consta de um único elemento v, diz-se que X é L.I., por definição, quando $v \neq 0$. Quando X é L.I., diz-se também que os elementos de X são *vetores linearmente independentes*.

Quando o conjunto X é L.I. seus elementos são todos $\neq 0$, pois o vetor nulo é combinação linear de quaisquer outros: $0 = 0 \cdot v_1 + \cdots + 0 \cdot v_m$. (Se não há "outros", $X = \{v\}$, $v \neq 0$.)

Um critério extremamente útil para verificar a independência linear de um conjunto é dado pelo teorema abaixo.

Teorema 3.1. *Seja X um conjunto* L.I. *no espaço vetorial E. Se $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$ com $v_1, \ldots, v_m \in X$ então $\alpha_1 = \cdots = \alpha_m = 0$. Reciprocamente, se a única combinação linear nula de vetores de X é aquela cujos coeficientes são todos iguais a zero, então X é um conjunto* L.I..

Demonstração: Suponhamos, por absurdo, que se tenha $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$ com $v_1, \dots, v_m \in X$ mas nem todos os α_i sejam nulos. Por simplicidade, seja $\alpha_1 \neq 0$. Então teremos $v_1 = -(\alpha_2/\alpha_1)v_2 - \cdots - (\alpha_m/\alpha_1)v_m = 0$, o que exprime v_1 como combinação linear de outros elementos de X. Reciprocamente, se X não fosse L.I., algum dos seus vetores seria combinação linear dos demais:

$$v = \alpha_1 v_1 + \cdots + \alpha_m v_m, \quad \text{logo} \quad 1 \cdot v - \alpha_1 v_1 - \cdots - \alpha_m v_m = 0,$$

uma combinação linear nula de vetores em X, na qual pelo menos o primeiro coeficiente não é zero. □

Corolário. *Se* $v = \alpha_1 v_1 + \cdots + \alpha_m v_m = \beta_1 v_1 + \cdots + \beta_m v_m$ *e os vetores* $v_1, \dots, v_m$ *são* L.I. *então* $\alpha_1 = \beta_1, \dots, \alpha_m = \beta_m$.

Com efeito, tem-se neste caso $(\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_m - \beta_m)v_m = 0$ logo $\alpha_1 - \beta_1 = \cdots = \alpha_m - \beta_m = 0$. □

Evidentemente, todo subconjunto de um conjunto L.I. é ainda L.I..

Exemplo 3.1. Os vetores canônicos $e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$ em $\mathbb{R}^n$ são L.I.. Com efeito, $\alpha_1 e_1 + \cdots + \alpha_n e_n = 0$ significa $(\alpha_1, \dots, \alpha_n) = 0$, logo $\alpha_1 = \cdots = \alpha_n = 0$. Analogamente, os monômios $1, x, \dots, x^n$ em $\mathcal{P}_n$ são L.I. pois $\alpha_o + \alpha_1 x + \cdots + \alpha_n x^n = p(x)$ é o vetor nulo em $\mathcal{P}_n$ somente quando $p(x)$ é a função identicamente nula, isto é, $p(x) = 0$ *para todo* $x \in \mathbb{R}$. Isto obriga a ser $\alpha_o = \cdots = \alpha_n = 0$ pois um polinômio não nulo de grau k tem no máximo k raízes reais. Esta observação nos permite ainda concluir que $X = \{1, x, \dots, x^n, \dots\} \subset \mathcal{P}$ é um conjunto infinito L.I..

Na prática, o critério seguinte é às vezes útil.

Teorema 3.2. Sejam $v_1, \dots, v_m$ vetores não-nulos do espaço vetorial E. Se nenhum deles é combinação linear dos anteriores então o conjunto $X = \{v_1, \dots, v_m\}$ é L.I..

Demonstração: Suponhamos, por absurdo, que uma combinação linear dos vetores dados, com coeficientes não todos nulos, fosse igual a zero. Se $\alpha_r v_r$ fosse a última parcela não-nula dessa combinação, teríamos então

$$\alpha_1 v_1 + \cdots + \alpha_r v_r = 0,$$

com $\alpha_r \neq 0$. Daí viria $v_r = -\frac{\alpha_1}{\alpha_r} v_1 - \cdots - \frac{\alpha_{r-1}}{\alpha_r} v_{r-1}$, logo v_r seria combinação linear dos elementos anteriores a ele na lista $v_1, \ldots, v_m$. (Observe que $r > 1$ pois $v_1 \neq 0$.) □

Observação: Evidentemente, vale um resultado análogo, com "subseqüentes" em vez de "anteriores" no enunciado.

Um conjunto $X \subset E$ diz-se *linearmente dependente* (abreviadamente, L.D.) quando não é L.I..

Isto significa que algum dos vetores $v \in X$ é combinação linear de outros elementos de X, ou então que $X = \{0\}$. A fim de que X seja L.D. é necessário e suficiente que exista uma combinação linear nula $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$ de vetores $v_1, \ldots, v_m \in X$ com algum coeficiente $\alpha_i \neq 0$. Se $X \subset Y$ e X é L.D. então Y também é L.D.. Se $0 \in X$ então o conjunto X é L.D..

Exemplo 3.2. Os vetores $u = (1, 2, 3)$, $v = (4, 5, 6)$, $w = (7, 8, 9)$ em $\mathbb{R}^3$ são L.D. pois $w = 2v - u$.

Exemplo 3.3. Quando os vetores $v_1, \ldots, v_m$ são L.D., isto não significa que qualquer um deles seja combinação linear dos demais. Por exemplo se $u = (1, 2)$, $v = (3, 4)$ e $w = (4, 8)$ então $\{u, v, w\} \subset \mathbb{R}^2$ é um conjunto L.D. pois $w = 4u + 0 \cdot v$ porém v não é combinação linear de u e w.

Uma *base* de um espaço vetorial E é um conjunto $\mathcal{B} \subset E$ linearmente independente que gera E. Isto significa que todo vetor $v \in E$ se exprime, de modo único, como combinação linear $v = \alpha_1 v_1 + \cdots + \alpha_m v_m$ de elementos $v_1, \ldots, v_m$ da base $\mathcal{B}$. Se $\mathcal{B} = \{v_1, \ldots, v_m\}$ é uma base de E e $v = \alpha_1 v_1 + \cdots + \alpha_m v_m$, então os números $\alpha_1, \ldots, \alpha_m$ chamam-se as *coordenadas* do vetor v na base $\mathcal{B}$.

Exemplo 3.4. Os vetores $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ constituem uma base $\{e_1, \ldots, e_n\}$ de $\mathbb{R}^n$, chamada a *base canônica*. Analogamente, os monômios $1, x, \ldots, x^n$ formam uma base para o espaço vetorial $\mathcal{P}_n$ dos polinômios de grau $\leq n$. O conjunto

$$\{1, x, \ldots, x^n, \ldots\}$$

dos monômios de graus arbitrários constitui uma base (infinita) para o espaço vetorial $\mathcal{P}$ de todos os polinômios reais. Convém observar, entretanto, que o conjunto $X = \{\bar{e}_1, \ldots, \bar{e}_n, \ldots\} \subset \mathbb{R}^\infty$, onde

$\bar{e}_n = (0, \ldots, 0, 1, 0, \ldots)$ é a seqüência infinita cujo n-ésimo termo é 1 e os demais são iguais a zero, é um conjunto infinito L.I. mas não é uma base de $\mathbb{R}^\infty$ pois não gera este espaço. Com efeito, o subespaço vetorial de $\mathbb{R}^\infty$ gerado por X é o conjunto $\mathbb{R}^{(\infty)}$ formado pelas seqüências $v = (\alpha_1, \ldots, \alpha_n, \ldots)$ nas quais apenas um número finito de coordenadas α_n é $\neq 0$.

Demonstraremos a seguir que se um espaço vetorial E admite uma base com n elementos então todas as bases de E têm o mesmo número n de elementos. Este número é chamado a *dimensão* de E.

O ponto de partida é o lema abaixo. Nele, um sistema linear é chamado *homogêneo* quando o segundo membro de cada equação é igual a zero. Todo sistema homogêneo admite pelo menos a *solução trivial* $(0, 0, \ldots, 0)$. Isto é coerente com o Exemplo 2.4, pois as soluções $v = (x_1, \ldots, x_n)$ de um sistema homogêneo constituem um subespaço vetorial de $\mathbb{R}^n$ e todo subespaço contém o vetor nulo.

Lema 3.1. *Todo sistema linear homogêneo cujo número de incógnitas é maior do que o número de equações admite uma solução não-trivial.*

Demonstração: Consideremos o sistema

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0, \end{aligned} \tag{*}$$

de m equações com n incógnitas, onde $m < n$. Usaremos indução no número m de equações. Para $m = 1$, temos uma única equação

$$a_{11}x_1 + \cdots + a_{1n}x_n = 0,$$

com $n > 1$ incógnitas. Um dos coeficientes a_{1i} é $\neq 0$. Mudando os nomes das incógnitas, se necessário, podemos supor que $a_{1n} \neq 0$. A equação dada equivale a

$$x_n = -\left(\frac{a_{11}}{a_{1n}}x_1 + \cdots + \frac{a_{1n-1}}{a_{1n}}x_{n-1}\right).$$

Atribuindo arbitrariamente valores não-nulos às $n-1$ incógnitas $x_1, \ldots, x_{n-1}$ e calculando x_n por meio desta última expressão, obtemos uma solução não-trivial $(x_1, \ldots, x_n)$ para a equação dada. Para

completar a indução, suponhamos o lema verdadeiro para um sistema com $m - 1$ equações. Mudando, se necessário, a ordem das equações e os nomes das incógnitas, podemos admitir que, no sistema (*) dado, tem-se $a_{mn} \neq 0$. Então da m-ésima equação resulta

$$x_n = -\left(\frac{a_{m1}}{a_{mn}} x_1 + \cdots + \frac{a_{mn-1}}{a_{mn}} x_{n-1}\right).$$

Substituindo, em cada uma das $m-1$ primeiras equações, a incógnita x_n por este valor, obtemos um sistema homogêneo de $m - 1$ equações nas $n - 1$ incógnitas $x_1, \ldots, x_{n-1}$. Pela hipótese de indução, este sistema admite uma solução não-trivial $(\alpha_1, \ldots, \alpha_{n-1})$, pois $n - 1 > m - 1$. Pondo

$$\alpha_n = -\left(\frac{a_{m1}}{a_{mn}} \alpha_1 + \cdots + \frac{a_{mn-1}}{a_{mn}} \alpha_{n-1}\right),$$

obtemos uma solução não-trivial $(\alpha_1, \ldots, \alpha_{n-1}, \alpha_n)$ do sistema proposto (*). □

Teorema 3.3. Se os vetores $v_1, \ldots, v_m$ geram o espaço vetorial E então qualquer conjunto com mais de m vetores em E é L.D.

Demonstração: Dados os vetores $w_1, \ldots, w_n$ em E, com $n > m$, para cada $j = 1, \ldots, n$ temos $w_j = \alpha_{1j}v_1 + \cdots + \alpha_{mj}v_m$ pois os vetores $v_1, \ldots, v_m$ geram E. Para mostrar que os vetores w_j são L.D., devemos achar coeficientes $x_1, \ldots, x_n$, não todos iguais a zero, tais que $x_1w_1 + \cdots + x_nw_n = 0$. Substituindo os w_j por suas expressões em termos dos v_i, esta igualdade significa que

$$\left(\sum_{j=1}^{n} x_j\alpha_{1j}\right) v_1 + \left(\sum_{j=1}^{n} x_j\alpha_{2j}\right) v_2 + \cdots + \left(\sum_{j=1}^{n} x_j\alpha_{mj}\right) v_m = 0.$$

Certamente esta última condição será satisfeita desde que todos os somatórios dentro dos parênteses sejam nulos, ou seja, que $(x_1, \ldots, x_n)$ seja uma solução não-trivial do sistema homogêneo

$$\begin{aligned}
\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= 0 \\
\alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n &= 0 \\
&\vdots \\
\alpha_{m1}x_1 + \alpha_{m2}x_2 + \cdots + \alpha_{mn}x_n &= 0.
\end{aligned}$$

Uma tal solução existe, pelo Lema 3.1, pois $n > m$. Logo $w_1, \ldots, w_n$ são L.D. e o teorema está demonstrado. □

Corolário 1. *Se os vetores* $v_1, \ldots, v_m$ *geram o espaço vetorial* E *e os vetores* $u_1, \ldots, u_n$ *são* L.I. *então* $n \leq m$.

Este corolário é uma mera reformulação do Teorema 3.3.

Corolário 2. *Se o espaço vetorial* E *admite uma base* $\mathcal{B}=\{u_1, \ldots, u_n\}$ *com* n *elementos, qualquer outra base de* E *possui também* n *elementos.*

Com efeito, seja $\mathcal{B}' = \{v_1, \ldots, v_m\}$ outra base de E. Como $\mathcal{B}'$ gera E e $\mathcal{B}$ é L.I., temos $n \leq m$, pelo Corolário 1. Como $\mathcal{B}$ gera E e $\mathcal{B}'$ é L.I., do mesmo corolário segue-se $m \leq n$. Logo $m = n$. □

Diz-se que o espaço vetorial E tem *dimensão finita* quando admite uma base $\mathcal{B} = \{v_1, \ldots, v_n\}$ com um número finito n de elementos. Este número, que é o mesmo para todas as bases de E, chama-se a *dimensão* do espaço vetorial E: $n = \dim E$. Por extensão, diz-se que o espaço vetorial $E = \{0\}$ tem dimensão zero.

Corolário 3. *Se a dimensão de* E *é* n*, um conjunto com* n *vetores gera* E *se, e somente se, é L.I..*

Com efeito, se $X = \{v_1, \ldots, v_n\}$ gera E e não é L.I. então um dos seus elementos é combinação dos $n-1$ restantes. Estes $n-1$ vetores formariam ainda um conjunto de geradores de E, em contradição com o Teorema 3.3, pois E contém (uma base com) n vetores linearmente independentes. Reciprocamente, suponhamos que X seja L.I.. Se X não gerasse E, existiria um vetor $v \in E$ que não seria combinação linear dos elementos de X. Então, pelo Teorema 3.2, $\{v_1, \ldots, v_n, v\}$ seria L.I., em contradição com o Teorema 3.3, pois uma base de E, com n elementos, gera o espaço. □

Como a base canônica $\{e_1, \ldots, e_n\} \subset \mathbb{R}^n$ tem n elementos, $\mathbb{R}^n$ é um espaço vetorial de dimensão finita n. Segue-se então do Corolário 3 que, para mostrarmos que n vetores $v_1, \ldots, v_n \in \mathbb{R}^n$ formam uma base basta provarmos que eles são L.I. ou, alternativamente, que geram $\mathbb{R}^n$.

Teorema 3.4. *Seja E um espaço vetorial de dimensão finita n. Então:*

(a) *Todo conjunto X de geradores de E contém uma base.*

(b) *Todo conjunto* L.I. $\{v_1, \ldots, v_m\} \subset E$ *está contido numa base.*

(c) *Todo subespaço vetorial $F \subset E$ tem dimensão finita, a qual é $\leq n$.*

(d) *Se a dimensão do subespaço $F \subset E$ é igual a n, então $F = E$.*

Demonstração: (a) Os conjuntos L.I. em E têm no máximo n elementos. Seja $Y = \{v_1, \ldots, v_m\} \subset X$ um subconjunto L.I. de X com o número máximo possível de elementos. Se existisse algum vetor $v \in X$ que não fosse combinação linear de $v_1, \ldots, v_m$ então o conjunto $\{v_1, \ldots, v_m, v\} \subset X$ seria L.I. pelo Teorema 3.2, mas isto contradiria a maximalidade de m. Logo devemos ter $X \subset S(Y)$, donde $E = S(X) \subset S(Y)$ e daí $S(Y) = E$, ou seja Y é uma base de E, contida em X, como se devia demonstrar.

(b) Seja

$$Y = \{v_1, \ldots, v_m, v_{m+1}, \ldots, v_k\}$$

um conjunto L.I. com o número máximo possível de elementos contendo os m vetores dados. (Pelo Teorema 3.3, tem-se $k \leq n$.) Se existisse em E algum vetor v que não fosse combinação linear dos elementos de Y, então $Y \cup \{v\}$ seria um conjunto L.I., de acordo com o Teorema 3.2, em contradição com a maximalidade de k. Segue-se que Y gera E, logo é uma base de E, contendo $v_1, \ldots, v_m$.

(c) Seja $Y = \{v_1, \ldots, v_m\} \subset F$ um subconjunto de F que é L.I. e tem o número máximo possível de elementos. Então Y gera F pois se algum elemento $v \in F$ não fosse combinação linear dos vetores de Y então, pelo Teorema 3.2, $\{v_1, \ldots, v_m, v\} \subset F$ seria um conjunto L.I., contrariando a maximalidade de m. Portanto Y é uma base de F e F tem dimensão finita. Além disso, tem-se $\dim F = m \leq n$ pois nenhum conjunto com mais de n elementos em E pode ser L.I. .

(d) Se $\dim F = \dim E = n$ então toda base de F é um subconjunto L.I. com n elementos em E, logo gera E, pelo Corolário 3. Segue-se que $F = E$. □

Neste livro, trataremos primordialmente dos espaços vetoriais de dimensão finita. Diz-se que o espaço vetorial E tem *dimensão infinita* quando ele não tem dimensão finita, isto é, quando nenhum subconjunto finito de E é uma base. Como todo subconjunto finito (que não se reduza ao vetor 0) contém um subconjunto L.I. que gera o mesmo subespaço, podemos dizer que um espaço vetorial E tem dimensão infinita se, e somente se, não é gerado por um conjunto finito de vetores. Por exemplo, o espaço $\mathcal{P}$ de todos os polinômios reais tem dimensão infinita pois se $X \subset \mathcal{P}$ é um conjunto finito de polinômios e r é o mais alto grau de um polinômio qualquer de X então o subespaço vetorial gerado por X está contido em $\mathcal{P}_r$, logo não é igual a $\mathcal{P}$.

Exemplo 3.5. Os monômios $1, x, \ldots, x^n$ constituem uma base do espaço vetorial $\mathcal{P}_n$, dos polinômios de grau $\leq n$, logo $\mathcal{P}_n$ tem dimensão finita e $\dim \mathcal{P}_n = n + 1$. Por outro lado, o conjunto infinito $\{1, x, \ldots, x^n, \ldots\}$ é uma base do espaço vetorial $\mathcal{P}$ de todos os polinômios, o qual tem dimensão infinita. Também tem dimensão infinita o espaço $\mathbb{R}^{(\infty)}$, introduzido no Exemplo 3.4, pois admite a base infinita $\{\overline{e}_1, \overline{e}_2, \ldots, \overline{e}_n, \ldots\}$, onde $\overline{e}_n = (0, \ldots, 0, 1, 0, \ldots)$ é a seqüência infinita cujo n-ésimo termo é 1 e os demais são zeros. Finalmente, embora não exibamos explicitamente uma base para o espaço $\mathbb{R}^\infty$, podemos assegurar que ele não tem dimensão finita, em virtude do item (c) do Teorema 3.4 acima, já que $\mathbb{R}^{(\infty)}$ é um subespaço de $\mathbb{R}^\infty$ com dimensão infinita.

Exemplo 3.6. O espaço vetorial $M(m \times n)$, das matrizes $m \times n$, tem dimensão finita, igual a $m \cdot n$. Uma base para $M(m \times n)$ é formada pelas matrizes $\mathbf{e}_{ij}$, cujo ij-ésimo elemento (na interseção da i-ésima linha com a j-ésima coluna) é igual a 1 e os demais elementos são iguais a zero.

Exemplo 3.7. Se os coeficientes $a_1, \ldots, a_n$ não são todos iguais a zero, o hiperplano

$$H = \{(x_1, \ldots, x_n) \in \mathbb{R}^n; a_1x_1 + \cdots + a_nx_n = 0\}$$

é um subespaço vetorial de dimensão $n - 1$ em $\mathbb{R}^n$. Com efeito, admitindo (por simplicidade) que $a_n \neq 0$, vemos que

$$v = (x_1, \ldots, x_n) \in H \Leftrightarrow x_n = -\frac{a_1}{a_n} x_1 - \cdots - \frac{a_{n-1}}{a_n} x_{n-1} .$$

Em particular, para todo $i = 1, \ldots, n-1$, o vetor

$$v_i = (0, \ldots, 1, \ldots, 0, -a_i/a_n),$$

cuja i-ésima coordenada é 1, a última é $-a_i/a_n$ e as demais são zero, pertence a H. Além disso, os vetores $v_1, \ldots, v_{n-1}$ são L.I., como se vê facilmente. Logo o subespaço H tem dimensão $n-1$ ou n. Como $H \neq \mathbb{R}^n$ (por exemplo, o vetor $v = (0, \ldots, 0, a_n)$ não pertence a H), segue-se que $\dim H = n-1$ e os vetores $v_1, \ldots, v_{n-1}$ formam uma base do hiperplano H.

Diz-se que a variedade afim $V \subset E$ tem *dimensão* r quando $V = x + F$, onde o subespaço vetorial $F \subset E$ tem dimensão r.

Exercícios

3.1. Dados os vetores $u = (a_1, a_2, a_3)$, $v = (b_1, b_2, b_3)$ e $w = (c_1, c_2, c_3)$, escreva $u' = (a_1, a_2)$, $v' = (b_1, b_2)$ e $w' = (c_1, c_2)$. Supondo u' e v' L.I., existem $\alpha, \beta \in \mathbb{R}$ tais que $w' = \alpha u' + \beta v'$. Prove que $\{u, v, w\}$ é L.D. se, e somente se, $w = \alpha u + \beta v$ (com os *mesmos* α e β). Use esse critério para determinar se os vetores u, v e w abaixo são L.I. ou L.D.:

(a) $u = (1, 2, 3), \quad v = (1, 3, 2), \quad w = (-1, 2, 3)$

(b) $u = (1, 2, 3), \quad v = (1, 3, 2), \quad w = (1, 4, 1).$

3.2. Mostre que as matrizes **a**, **b** e **c** abaixo são L.I.:

$$\mathbf{a} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

3.3. Prove que os polinômios seguintes são linearmente independentes:

$$p(x) = x^3 - 5x^2 + 1, \quad q(x) = 2x^4 + 5x - 6, \quad r(x) = x^2 - 5x + 2.$$

3.4. Seja X um conjunto de polinômios. Se dois polinômios quaisquer pertencentes a X têm graus diferentes, prove que X é L.I. .

3.5. No espaço $\mathcal{P}_3$ dos polinômios de grau ≤ 3, verifique se os polinômios abaixo são L.I. ou L.D.:

$$p(x) = x^3 - 3x^2 + 5x + 1,$$
$$q(x) = x^3 - x^2 + 6x + 2,$$
$$r(x) = x^3 - 7x^2 + 4x.$$

3.6. Se uma função em $C^\infty(\mathbb{R})$ é combinação linear de outras então suas derivadas sucessivas são combinações lineares (com os mesmos coeficientes) das derivadas dessas outras. Use este fato para mostrar que $\{e^x, e^{2x}, x^3, x^2, x\}$ é um conjunto L.I. .

3.7. Seja $E = F_1 \oplus F_2$. Se $\mathcal{B}_1$ é uma base de F_1 e $\mathcal{B}_2$ é uma base de F_2 , prove que $\mathcal{B}_1 \cup \mathcal{B}_2$ é uma base de E.

3.8. Exiba uma base para cada um dos subespaços de $\mathbb{R}^4$ listados a seguir:

$$F = \{(x_1, x_2, x_3, x_4); x_1 = x_2 = x_3 = x_4\}$$
$$G = \{(x_1, x_2, x_3, x_4); x_1 = x_2 \text{ e } x_3 = x_4\}$$
$$H = \{(x_1, x_2, x_3, x_4); x_1 = x_2 = x_3\}$$
$$K = \{(x_1, x_2, x_3, x_4); x_1 + x_2 + x_3 + x_4 = 0\}.$$

3.9. Seja E um espaço vetorial de dimensão finita. Dado um subespaço $F \subset E$, prove que se pode obter um subespaço $G \subset E$ tal que $E = F \oplus G$.

3.10. Seja F o subespaço vetorial (plano) de $\mathbb{R}^3$ formado pelos vetores $v = (x, y, z)$ tais que $x - 2y + 4z = 0$. Obtenha uma base $\{u_1, u_2, u_3\} \subset \mathbb{R}^3$ tal que u_1 e u_2 pertençam a F.

3.11. Mostre que os polinômios 1, $x-1$ e x^2-3x+1 formam uma base de $\mathcal{P}_2$. Exprima o polinômio $2x^2 - 5x + 6$ como combinação linear dos elementos dessa base.

3.12. Mostre que os vetores $u = (1, 1)$ e $v = (-1, 1)$ formam uma base de $\mathbb{R}^2$. Exprima cada um dos vetores $e_1 = (1, 0)$ e $e_2 = (0, 1)$ como combinação linear dos elementos dessa base.

3.13. Mostre que os vetores $u = (1, 1, 1)$, $v = (1, 2, 1)$ e $w = (2, 1, 2)$ são L.D. .

3.14. Assinale V(erdadeiro) ou F(also) quanto à validez da afirmação:
"A união de dois subconjuntos L.I. do espaço vetorial E é ainda um conjunto L.I."

( ) Sempre.

( ) Nunca.

( ) Quando um deles é disjunto do outro.

( ) Quando um deles é parte do outro.

( ) Quando um deles é disjunto do subespaço gerado pelo outro.

( ) Quando o número de elementos de um deles mais o número de elementos do outro é igual à dimensão de E.

3.15. Seja S o conjunto das matrizes simétricas $n \times n$. Para cada par (i, j) de números naturais de 1 até n, com $i \leq j$, seja $\mathbf{s}_{ij}$ a matriz $n \times n$ cujos elementos nas posições ij e ji são iguais a 1 e os demais são zero. Prove que estas matrizes constituem uma base para o subespaço vetorial $S \subset M(n \times n)$. De modo análogo, obtenha uma base do subespaço A das matrizes anti-simétricas $n \times n$. Conclua que dim $S = n(n+1)/2$ e dim $A = n(n-1)/2$.

3.16. As matrizes $\mathbf{t} = [t_{ij}] \in M(n \times n)$ tais que $t_{ij} = 0$ quando $i < j$ são chamadas *triangulares inferiores*. Prove que elas constituem um subespaço vetorial $L \subset M(n \times n)$, obtenha uma base para L e determine a sua dimensão.

3.17. Obtenha uma base e conseqüentemente determine a dimensão de cada um dos subespaços de $M(n \times n)$ abaixo descritos:

(a) matrizes cuja soma dos elementos da diagonal (traço) é zero.

(b) matrizes que têm a primeira e a última linha iguais.

(c) matrizes cuja segunda linha é igual à terceira coluna.

(d) matrizes nas quais a soma dos elementos da primeira linha é igual à soma dos elementos da segunda coluna.

3.18. Sejam $u, v \in E$ vetores linearmente independentes. Dado $\alpha \neq 0$, prove que o conjunto de dois elementos $\{v, v + \alpha u\}$ é uma base do subespaço gerado pelos vetores $v, v + u, v + 2u, \ldots, v + nu, \ldots$.

3.19. Sejam $v_1 = (1, 2, \ldots, n)$, $v_2 = (n + 1, n + 2, \ldots, 2n), \ldots, v_n = (n^2 - n + 1, n^2 - n + 2, \ldots, n^2)$. Prove que estes vetores geram em $\mathbb{R}^n$ o mesmo subespaço F que os vetores $w_1 = (1, n+1, 2n+1, \ldots, n^2 - n + 1)$, $w_2 = (2, n+2, \ldots, n^2 - n + 2), \ldots, w_n = (n, 2n, \ldots, n^2)$ e que $\dim F = 2$. (Veja Exercício 2.3.)

3.20. Ache uma solução não-trivial para o sistema homogêneo:

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & 3x_3 & + & 4x_4 & = & 0 \\ 2x_1 & + & x_2 & + & x_3 & - & x_4 & = & 0 \\ 3x_1 & - & 2x_2 & + & x_3 & - & 2x_4 & = & 0 \end{array}$$

e, a partir daí, obtenha uma combinação linear nula dos vetores $v_1 = (1, 2, 3)$, $v_2 = (2, 1, -2)$, $v_3 = (3, 1, 1)$, $v_4 = (4, -1, -2)$, na qual os coeficientes não são todos iguais a zero.

3.21. Seja $\{v_1, \ldots, v_n\}$ uma base do espaço vetorial E. Se os números $a_1, \ldots, a_n$ não são todos iguais a zero, prove que o conjunto F dos vetores $v = x_1 v_1 + \cdots + x_n v_n$ tais que $a_1 x_1 + \cdots + a_n x_n = 0$ é um subespaço vetorial de E, com $\dim F = n - 1$.

3.22. Prove que $\{1, e^x, e^{2x}, e^{3x}, e^{4x}\}$ é um conjunto L.I. no espaço $C^\infty(\mathbb{R})$. (Sugestão: dada uma combinação linear nula, derive-a, depois divida por e^x e prossiga.)

3.23. Sejam $X_1, \ldots, X_n, \ldots$ subconjuntos L.I. do espaço vetorial E.

(a) Se $X_1 \subset X_2 \subset \ldots \subset X_n \subset X_{n+1} \subset \ldots$, prove que $X = \bigcup X_n$ é L.I..

(b) Se cada X_n tem n elementos, prove que existe um conjunto linearmente independente $X^* = \{x_1, \ldots, x_n, \ldots\}$ com $x_n \in X_n$ para cada $n \in \mathbf{N}$.

(c) Supondo $E = \mathbb{R}^{(\infty)}$ e admitindo as hipóteses dos ítens anteriores, é verdade que $X = \bigcup X_n$ seja uma base de E?

3.24. Se os vetores $v_1, \ldots, v_m$ são L.I., prove que o mesmo se dá com os vetores $v_1, v_2 - v_1, \ldots, v_m - v_1$. Vale a recíproca?

3.25. Dado o conjunto finito $X = \{a_1, \ldots, a_n\}$, obtenha uma base para o espaço vetorial $\mathcal{F}(X; \mathbb{R})$.

3.26. Seja X um conjunto infinito. Para cada $a \in X$, seja $f_a \colon X \to \mathbb{R}$ a função tal que $f_a(a) = 1$ e $f_a(x) = 0$ se $x \neq a$. Prove que o conjunto $Y \subset \mathcal{F}(X;\mathbb{R})$ formado por estas funções é linearmente independente, logo $\mathcal{F}(X;\mathbb{R})$ não tem dimensão finita. Prove ainda que Y não gera $\mathcal{F}(X;\mathbb{R})$.

3.27. Sejam $F_1, F_2 \subset E$ subespaços de dimensão finita. Obtenha uma base do subespaço $F_1 + F_2$ que contenha uma base de F_1, uma base de F_2 e uma base de $F_1 \cap F_2$.

3.28. Exiba uma base para cada um dos espaços vetoriais abaixo e daí calcule sua dimensão:

(a) polinômios pares de grau $\leq n$.

(b) polinômios ímpares de grau $\leq n$.

(c) polinômios de grau $\leq n$ que se anulam para $x = 2$ e $x = 3$.

(d) vetores de $\mathbb{R}^n$ $(n \geq 6)$ nos quais a segunda, a quarta e a sexta coordenadas são iguais.

3.29. Pode-se ter uma base de $\mathcal{P}_n$ formada por $n + 1$ polinômios de grau n?

3.30. Mostre que os vetores $u = (1,1,1)$, $v = (1,2,3)$ e $w = (1,4,9)$ formam uma base de $\mathbb{R}^3$. Exprima cada um dos vetores e_1, e_2, e_3 da base canônica de $\mathbb{R}^3$ como combinação linear de u, v e w.

3.31. Ache uma seqüência infinita $F_1, F_2, \ldots, F_n, \ldots$ de subespaços vetoriais de $\mathcal{P}$ tais que: (a) $\dim F_n = \infty$; (b) $F_m \cap F_n = \{0\}$ se $m \neq n$.

3.32. Para $1 \leq i \leq m$ e $1 \leq j \leq n$, sejam $s_i, t_j \colon M(m \times n) \to \mathbb{R}$ as funções definidas por $s_i(\mathbf{a})$ = soma dos elementos da i-ésima linha de $\mathbf{a}$ e $t_j(\mathbf{a})$ = soma dos elementos da j-ésima coluna de $\mathbf{a}$. Prove que $s_1, \ldots, s_m, t_1, \ldots, t_n$ são L.D. no espaço vetorial $E = \mathcal{F}(M(m \times n);\mathbb{R})$ mas o conjunto $\{s_1, \ldots, s_{m-1}, t_1, \ldots, t_n\}$ é L.I..

3.33. Com as notações do exercício anterior, sejam $\tau, \sigma \colon M(n \times n) \to \mathbb{R}$ as funções definidas, para cada $\mathbf{a} = [a_{ij}] \in M(n \times n)$ por $\tau(\mathbf{a}) = a_{11} + \cdots + a_{nn}$ (soma dos termos da diagonal principal) e $\sigma(\mathbf{a}) = a_{1n} + a_{2,n-1} + \cdots + a_{n1}$ (soma dos termos da outra diagonal). Prove

que, para $n \geq 3$, $\{s_1, \ldots, s_{n-1}, t_1, \ldots, t_n, \tau, \sigma\}$ são funções linearmente independentes.

3.34. Num espaço vetorial E, diz-se que o vetor v é uma *combinação afim* dos vetores $v_1, \ldots, v_r$ quando se tem $v = \alpha_1 v_1 + \cdots + \alpha_r v_r$, com $\alpha_1 + \cdots + \alpha_r = 1$. Diz-se que os vetores $v_1, \ldots, v_r$ são *afim-independentes* quando nenhum deles é uma combinação afim dos demais. Prove que as seguintes afirmações são equivalentes:

(1) Os vetores $v_1, \ldots, v_r$ são afim-independentes.

(2) Se $\alpha_1 v_1 + \cdots + \alpha_r v_r = 0$ e $\alpha_1 + \cdots + \alpha_r = 0$ então $\alpha_1 = \cdots = \alpha_r = 0$.

(3) Se $\alpha_1 v_1 + \cdots + \alpha_r v_r = \beta_1 v_1 + \cdots + \beta_r v_r$ com $\sum_{i=1}^{r} \alpha_i = \sum_{i=1}^{r} \beta_i$ então $\alpha_1 = \beta_1, \ldots, \alpha_r = \beta_r$. (Em particular, duas combinações afins dos v_i só podem ser iguais quando tiverem os mesmos coeficientes.)

(4) Os vetores $v_2 - v_1, v_3 - v_1, \ldots, v_r - v_1$ são L.I. .

(5) A variedade afim gerada por $v_1, \ldots, v_r$ tem dimensão $r - 1$.

4

Transformações Lineares

A Álgebra Linear pode ser apresentada sob três pontos de vista equivalentes: transformações lineares, matrizes ou formas quadráticas. A ênfase (ou até mesmo a exclusividade) que se dá a uma dessas abordagens é muitas vezes uma questão de hábito, gosto pessoal ou convicção. Neste livro, os três aspectos serão devidamente tratados porém a primazia será concedida às transformações lineares, pelos três motivos apontados, principalmente o último.

Sejam E, F espaços vetoriais. Uma *transformação linear* $A\colon E \to F$ é uma correspondência que associa a cada vetor $v \in E$ um vetor $A(v) = A \cdot v = Av \in F$ de modo que valham, para quaisquer $u, v \in E$ e $\alpha \in \mathbb{R}$, as relações:

$$A(u + v) = Au + Av,$$
$$A(\alpha \cdot v) = \alpha \cdot Av.$$

O vetor $A \cdot v$ chama-se a *imagem* (ou o *transformado*) de v pela transformação A.

Se $A\colon E \to F$ é uma transformação linear então $A \cdot 0 = 0$. Com efeito, $A \cdot 0 = A(0 + 0) = A \cdot 0 + A \cdot 0$. Além disso, dados $u, v \in E$ e $\alpha, \beta \in \mathbb{R}$, tem-se $A(\alpha u + \beta v) = A(\alpha u) + A(\beta v) = \alpha \cdot Au + \beta \cdot Av$. Mais geralmente, dados $v_1, \ldots, v_m$ em E e $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$, vale

$$A(\alpha_1 v_1 + \cdots + \alpha_m v_m) = \alpha_1 \cdot Av_1 + \cdots + \alpha_m \cdot Av_m.$$

Daí resultam $A(-v) = -Av$ e $A(u - v) = Au - Av$.

A *soma* de duas transformações lineares $A, B \colon E \to F$ e o *produto* de uma transformação linear $A \colon E \to F$ por um número $\alpha \in \mathbb{R}$ são as transformações lineares $A + B \colon E \to F$ e $\alpha A \colon E \to F$, definidas respectivamente por $(A + B)v = Av + Bv$ e $(\alpha A)v = \alpha \cdot Av$, para todo $v \in E$. O símbolo 0 indica a transformação linear nula $0 \colon E \to F$, definida por $0 \cdot v = 0$ e, definindo $-A \colon E \to F$ por $(-A) \cdot v = -Av$, vê-se que $(-A) + A = A + (-A) = 0$.

Seja $\mathcal{L}(E; F)$ o conjunto das transformações lineares de E em F. As definições acima tornam $\mathcal{L}(E; F)$ um espaço vetorial. Quando $E = F$, usaremos a notação $\mathcal{L}(E)$ em vez de $\mathcal{L}(E; E)$. As transformações lineares $A \colon E \to E$ do espaço vetorial E em si mesmo são chamadas *operadores lineares* em E. Por sua vez, as transformações lineares $\varphi \colon E \to \mathbb{R}$, com valores numéricos, são chamadas *funcionais lineares*. Escreve-se E^* em vez de $\mathcal{L}(E; \mathbb{R})$ e o conjunto E^* dos funcionais lineares $\varphi \colon E \to \mathbb{R}$ chama-se o espaço vetorial *dual* de E.

Um operador linear especial é o operador *identidade* $I \colon E \to E$, definido por $I \cdot v = v$ para todo $v \in E$. Quando for necessário especificar, escreveremos I_E em vez de I.

Uma transformação linear $A \colon E \to F$ é um tipo particular de função que tem o espaço vetorial E como domínio e o espaço F como contra-domínio. Em geral, para se definir uma função $f \colon X \to Y$ é necessário especificar o valor $f(x)$ para cada elemento x no seu domínio X. O que torna as transformações lineares tão manejáveis é que, para se conhecer $A \in \mathcal{L}(E; F)$, basta que se saibam os valores $A \cdot v$ que A assume nos vetores $v \in \mathcal{B}$, onde $\mathcal{B}$ é uma base de E. Isto é particularmente útil quando E tem dimensão finita. Neste caso, um número finito de valores $A \cdot v_1, \ldots, A \cdot v_n$ (onde $\{v_1, \ldots, v_n\} \subset E$ é uma base) atribuídos arbitrariamente, definem inteiramente uma transformação linear $A \colon E \to F$. Mais precisamente, vale o

Teorema 4.1. Sejam E, F espaços vetoriais e $\mathcal{B}$ uma base de E. A cada vetor $u \in \mathcal{B}$, façamos corresponder (de maneira arbitrária) um vetor $u' \in F$. Então existe uma única transformação linear $A \colon E \to F$ tal que $A \cdot u = u'$ para cada $u \in \mathcal{B}$.

Demonstração: Todo vetor $v \in E$ se exprime, de modo único, como uma combinação linear $v = \alpha_1 u_1 + \cdots + \alpha_m u_m$ de elementos $u_1, \ldots$ $\ldots, u_m$ da base $\mathcal{B}$. Definimos $A \colon E \to F$ pondo

$$A \cdot v = \alpha_1 u_1' + \cdots + \alpha_m u_m'.$$

Dados $v, w \in E$ temos

$$v = \alpha_1 u_1 + \cdots + \alpha_m u_m$$

e

$$w = \beta_1 u_1 + \cdots + \beta_m u_m\,.$$

(Mesmo que a base $\mathcal{B}$ seja infinita, podemos exprimir v e w como combinações lineares dos mesmos elementos de $\mathcal{B}$, completando com coeficientes zero os múltiplos dos u_i que aparecem apenas numa das duas expressões.) Então

$$v + w = \sum_{i=1}^{m} (\alpha_i + \beta_i) u_i$$

logo

$$A(v + w) = \Sigma(\alpha_i + \beta_i)u_i' = \Sigma\alpha_i u_i' + \Sigma\beta_i u_i' = A \cdot v + A \cdot w.$$

De maneira análoga se vê que $A(\alpha v) = \alpha \cdot Av$, portanto $A\colon E \to F$, assim definida, é uma transformação linear, tal que $A \cdot u = u'$, para todo $u \in \mathcal{B}$. Quanto à unicidade, seja $B\colon E \to F$ outra transformação linear tal que $B \cdot u = u'$ para todo $u \in \mathcal{B}$. Então, para cada $v = \Sigma\alpha_i u_i \in E$ tem-se

$$B \cdot v = B(\Sigma\alpha_i u_i) = \Sigma\alpha_i \cdot Bu_i = \Sigma\alpha_i \cdot u_i' = A \cdot v$$

portanto $B = A$. Isto completa a demonstração. □

Em virtude do Teorema 4.1, se quisermos definir uma transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ basta escolher, para cada $j = 1, \ldots, n$, um vetor $v_j = (a_{1j}, a_{2j}, \ldots, a_{mj}) \in \mathbb{R}^m$ e dizer que $v_j = A \cdot e_j$ é a imagem do j-ésimo vetor da base canônica, $e_j = (0, \ldots, 1, \ldots, 0)$, pela transformação linear A. A partir daí, fica determinada a imagem $A \cdot v$ de qualquer vetor $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Com efeito, tem-se $v = x_1 e_1 + \cdots + x_n e_n$, logo

$$A \cdot v = A\left(\sum_{j=1}^{n} x_j e_j\right) = \sum_{j=1}^{n} x_j A \cdot e_j = \sum_{j=1}^{n} (a_{1j}x_j, a_{2j}x_j, \ldots, a_{mj}x_j)$$

$$= \left(\sum_{j=1}^{n} a_{1j}x_j, \sum_{j=1}^{n} a_{2j}x_j, \ldots, \sum_{j=1}^{n} a_{mj}x_j\right),$$

ou seja,

$$A(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_m),$$

onde

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n\,. \end{aligned} \tag{*}$$

Resumindo: uma transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ fica inteiramente determinada por uma matriz $\mathbf{a} = [a_{ij}] \in M(m \times n)$. Os vetores-coluna dessa matriz são as imagens $A \cdot e_j$ dos vetores da base canônica de $\mathbb{R}^n$. A imagem $A \cdot v$ de um vetor arbitrário $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$ é o vetor $w = (y_1, \ldots, y_m) \in \mathbb{R}^m$ cujas coordenadas são dadas pelas equações (*) acima, nas quais ocorrem os vetores-linha da matriz $\mathbf{a}$. Diz-se que $\mathbf{a}$ é a *matriz da transformação* A relativa às bases canônicas de $\mathbb{R}^n$ e $\mathbb{R}^m$. Tem-se

$$A \cdot e_j = \sum_{i=1}^{m} a_{ij} e_i \quad (j = 1, \ldots, n),$$

onde os e_j estão em $\mathbb{R}^n$ e os e_i em $\mathbb{R}^m$.

Em particular, a matriz de um funcional linear $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ é do tipo $1 \times n$, logo pode ser escrita simplesmente como $[a_1, a_2, \ldots, a_n]$, onde $a_j = \varphi(e_j)$. Para todo vetor $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ tem-se $\varphi(x) = a_1x_1 + \cdots + a_nx_n$.

Na situação dual, uma transformação linear $A\colon \mathbb{R} \to \mathbb{R}^n$ é dada por uma matriz $n \times 1$, cuja única coluna é o vetor $v = A \cdot 1 = (a_1, \ldots, a_n)$. (A base canônica de $\mathbb{R}^1 = \mathbb{R}$ tem um único elemento $e_1 = 1$.) Assim, a transformação linear $A\colon \mathbb{R} \to \mathbb{R}^n$ fica inteiramente determinada por um único vetor $v \in \mathbb{R}^n$. Tem-se $A \cdot t = t \cdot v$ para todo $t \in \mathbb{R}$. Evidentemente, o mesmo se pode dizer de toda transformação linear $A\colon \mathbb{R} \to F$, seja qual for o espaço vetorial F: conhecendo $v = A \cdot 1 \in F$ tem-se $A \cdot t = tv$ para todo $t \in \mathbb{R}$.

Exemplo 4.1. Se $\dim E = 1$, todo operador $A\colon E \to E$ é do tipo $A = \alpha I$, isto é, existe uma constante $\alpha \in \mathbb{R}$ tal que $Av = \alpha v$ para

todo $v \in E$. Com efeito, seja $u \in E$ um vetor não-nulo. Então $\{u\} \subset E$ é uma base: todo vetor em E é múltiplo de u. Portanto existe $\alpha \in \mathbb{R}$ tal que $Au = \alpha u$. Para qualquer outro vetor $v \in E$, temos $v = \lambda u$ portanto $Av = A(\lambda u) = \lambda\, Au = \lambda\, \alpha u = \alpha(\lambda\, u) = \alpha v$.

Exemplo 4.2. (Rotação de ângulo θ em torno da origem em $\mathbb{R}^2$.) Trata-se do operador $R\colon \mathbb{R}^2 \to \mathbb{R}^2$, que leva cada vetor v no vetor Rv que dele resulta pela rotação de ângulo θ em torno da origem. A Fig. 4.1 deixa claro que $R(u+v) = R\cdot u + R\cdot v$. É bem mais claro ainda que $R(\alpha v) = \alpha \cdot Rv$ para $v \in \mathbb{R}^2$ e $\alpha \in \mathbb{R}$, logo R é uma transformação linear. Para um vetor $v = (x, y) \in \mathbb{R}^2$ arbitrário, seja $R \cdot v = (x', y')$. Sabemos que $x' = ax + by$ e $y' = cx + dy$ e

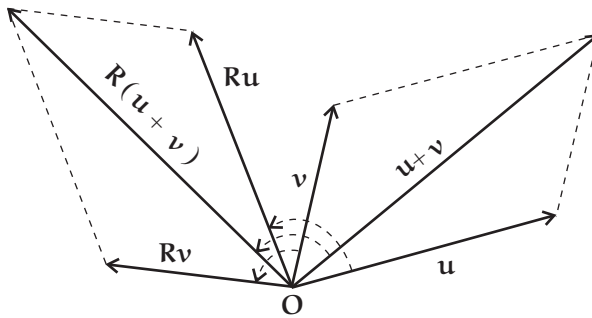



Figura 4.1 – Rotação de vetores.

queremos determinar a matriz

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

onde $Re_1 = (a, c)$ e $Re_2 = (b, d)$, com $e_1 = (1, 0)$ e $e_2 = (0, 1)$.

Ora, pelas definições de seno e cosseno, o vetor unitário Re_1, que forma com e_1 um ângulo θ, tem coordenadas $\cos\theta$ e $\operatorname{sen}\theta$, ou seja, $Re_1 = (\cos\theta, \operatorname{sen}\theta)$. Além disso, como e_2 forma com e_1 um ângulo reto, Re_2 também forma com Re_1 um ângulo reto. Logo $Re_2 = (-\operatorname{sen}\theta, \cos\theta)$. (Veja Fig. 4.2.)

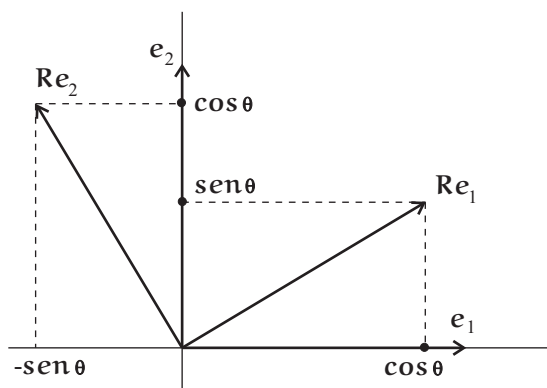



Figura 4.2 – Rotação de ângulo θ.

Portanto, a rotação $R: \mathbb{R}^2 \to \mathbb{R}^2$ leva um vetor $v = (x, y)$ no vetor $Rv = (x', y')$, onde

$$x' = x\cos\theta - y\,\mathrm{sen}\,\theta;$$
$$y' = x\,\mathrm{sen}\,\theta + y\cos\theta.$$

A matriz de R relativa à base canônica de $\mathbb{R}^2$ é

$$\begin{bmatrix} \cos\theta & -\mathrm{sen}\,\theta \\ \mathrm{sen}\,\theta & \cos\theta \end{bmatrix}.$$

Exemplo 4.3. (Projeção ortogonal sobre uma reta.) A reta $y = ax$ é o conjunto dos pontos $(x, ax) \in \mathbb{R}^2$, onde x varia em $\mathbb{R}$. Ela é o subespaço vetorial de $\mathbb{R}^2$ gerado pelo vetor $(1, a)$. Consideremos o operador $P: \mathbb{R}^2 \to \mathbb{R}^2$ que faz corresponder a cada $v = (x, y) \in \mathbb{R}^2$ o vetor $Pv = (x', ax')$, cuja extremidade é o pé da perpendicular baixada de v sobre a reta $y = ax$. (Veja Fig. 4.3.)

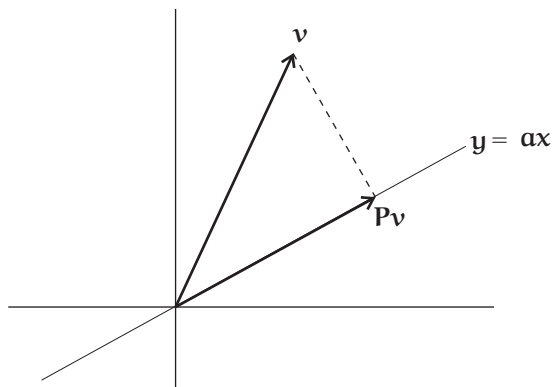



Figura 4.3 – Projeção ortogonal sobre uma reta.

Queremos determinar x' em função de x e y, o que nos dará as coordenadas (x', ax') de Pv em função das coordenadas de v. No caso particular em que $a = 0$, a reta $y = ax$ é o eixo das abcissas e a projeção Pv é simplesmente igual a $(x, 0)$. As equações da projeção P sobre o eixo horizontal são portanto $x' = x$, $y' = 0$. A matriz de P na base canônica de $\mathbb{R}^2$ é $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. No caso geral, a extremidade do vetor Pv é o vértice do ângulo reto num triângulo retângulo cujos demais vértices são a origem e a extremidade do vetor v. Pelo teorema de Pitágoras, temos

$$\text{dist}(v, 0)^2 = \text{dist}(Pv, 0)^2 + \text{dist}(v, Pv)^2,$$

ou seja,

$$x^2 + y^2 = (x')^2 + a^2(x')^2 + (x - x')^2 + (y - ax')^2.$$

Suponhamos $x' \neq 0$. Desenvolvendo, simplificando e dividindo ambos os membros por x', obtemos $(1 + a^2)x' = x + ay$, donde

$$x' = \frac{x + ay}{1 + a^2}, \quad \text{ou seja} \quad x' = \frac{1}{1 + a^2}\, x + \frac{a}{1 + a^2}\, y.$$

O caso $x' = 0$ significa que $v = (x, y)$ está sobre a perpendicular à reta $y = ax$ passando pela origem. Ora, a equação dessa perpendicular é $x + ay = 0$, logo a expressão $x' = (x + ay)/(1 + a^2)$ fornece x' em função de x e y em todos os casos. Vemos, em particular, que

a projeção $P\colon \mathbb{R}^2 \to \mathbb{R}^2$ é um operador linear, cuja matriz na base canônica de $\mathbb{R}^2$ é

$$\begin{bmatrix} \frac{1}{1+a^2} & \frac{a}{1+a^2} \\ \\ \frac{a}{1+a^2} & \frac{a^2}{1+a^2} \end{bmatrix}.$$

Exemplo 4.4. (Reflexão em torno de uma reta.) Seja $S\colon \mathbb{R}^2 \to \mathbb{R}^2$ a reflexão em torno da reta $y = ax$. Para todo $v = (x, y) \in \mathbb{R}^2$, a reta $y = ax$ é a bissetriz do ângulo entre v e Sv e é perpendicular à reta que liga v a Sv. Seja $P\colon \mathbb{R}^2 \to \mathbb{R}^2$ a projeção ortogonal sobre a reta $y = ax$. A Fig. 4.4 mostra que, para todo $v \in \mathbb{R}^2$, tem-se $v + Sv = 2Pv$, ou seja, que $I + S = 2P$, onde $I\colon \mathbb{R}^2 \to \mathbb{R}^2$ é o operador identidade. Daí vem $S = 2P - I$. Usando o exemplo anterior, concluímos que, para todo $v = (x, y)$, tem-se $Sv = (x', y')$, onde

$$x' = \frac{1-a^2}{1+a^2}\,x + \frac{2a}{1+a^2}\,y, \qquad y' = \frac{2a}{1+a^2}\,x - \frac{1-a^2}{1+a^2}\,y\,.$$

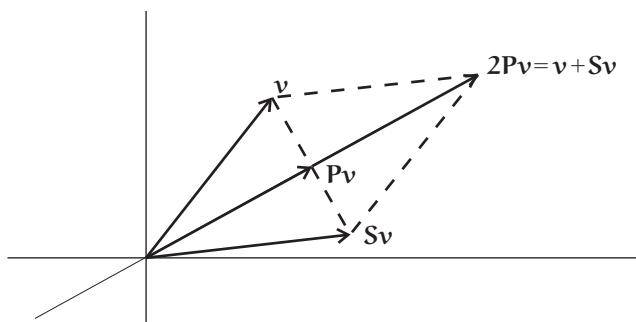



Figura 4.4 – Reflexão em torno de uma reta.

Exemplo 4.5. Como vimos acima, o único tipo de funcional linear $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ é o da forma $\varphi(v)=a_1x_1 + \cdots + a_nx_n$, para $v = (x_1, \ldots, x_n)$. Por outro lado, se $E = C^o([a, b])$ é o espaço vetorial das funções contínuas $f\colon [a, b] \to \mathbb{R}$, podemos definir o funcional linear $\varphi\colon E \to \mathbb{R}$ pondo

$$\varphi(f) = \int_a^b f(x)\,dx.$$

Outro exemplo de funcional linear ξ em E consiste em fixar um ponto $c \in [a, b]$ e definir, para cada $f \in E$, $\xi(f) = f(c)$. Ainda no contexto

do espaço de funções $E = C^o([a,b])$, podemos definir um operador linear $K\colon E \to E$ do seguinte modo: fixamos uma função contínua $k\colon [a,b] \times [a,b] \to \mathbb{R}$, de duas variáveis, e fazemos corresponder a cada $f \in E$ a função $g = Kf \in E$ dada por

$$g(x) = \int_a^b k(x,y) f(y)\, dy.$$

Finalmente, temos o importante operador de derivação $D\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$, definido por $Df = f' =$ derivada de f. Ele também pode ser considerado, de forma mais restrita, como um operador $D\colon \mathcal{P}_n \to \mathcal{P}_n$, onde

$$D\left(\sum_{i=0}^{n} a_i x^i\right) = \sum_{i=1}^{n} i a_i x^{i-1},$$

ou, de forma mais ampla, como uma transformação linear $D\colon C^k(\mathbb{R}) \to C^{k-1}(\mathbb{R})$, para $k > 0$.

Exercícios

4.1. Prove que se $A, B\colon E \to F$ são transformações lineares e α é um número real então $A + B$ e αA, conforme definidas no texto, são transformações lineares.

4.2. Sejam $R, P, S\colon \mathbb{R}^2 \to \mathbb{R}^2$ respectivamente a rotação de $30°$ em torno da origem, a projeção ortogonal sobre a reta $y = x/3$ e a reflexão em torno da mesma reta. Dado o vetor $v = (2,5)$, determine os vetores Rv, Pv e Sv.

4.3. Assinale verdadeiro (V) ou falso (F):
É dada uma transformação linear $A\colon E \to F$.

( ) Se $v \in E$ é tal que $Av = 0$ então $v = 0$.

( ) Se $Aw = Au + Av$ então $w = u + v$.

( ) Se v é combinação linear de $u_1, \ldots, u_m$ então Av é combinação linear de $Au_1, \ldots, Au_m$.

( ) Se $u, v, w \in E$ são colineares (isto é, pertencentes a uma mesma reta) então Au, Av e Aw são colineares.

4.4. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ a projeção sobre o eixo x, paralelamente à reta $y = ax$ $(a \neq 0)$. Isto significa que, para todo $v = (x, y)$, tem-se $Av = (x', 0)$, tal que $v - Av$ pertence à reta $y = ax$. Exprima x' em função de x e y e escreva a matriz de A relativamente à base canônica de $\mathbb{R}^2$.

4.5. Dados os vetores $u_1 = (2, -1)$, $u_2 = (1, 1)$, $u_3 = (-1, -4)$, $v_1 = (1, 3)$, $v_2 = (2, 3)$ e $v_3 = (-5, -6)$, decida se existe ou não um operador linear $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ tal que $Au_1 = v_1$, $Au_2 = v_2$ e $Au_3 = v_3$. Mesma pergunta com $v_3 = (5, -6)$ e com $v_3 = (5, 6)$.

4.6. A expressão geral de um operador linear $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ é $A(x, y) = (ax + by, cx + dy)$. Determine as constantes a, b, c e d de modo que A transforme os vetores $u = (1, 2)$ e $v = (3, 4)$ nos vetores $Au = (1, 1)$ e $Av = (2, 2)$.

4.7. A expressão geral de um funcional linear $f\colon \mathbb{R}^3 \to \mathbb{R}$ é $f(x, y, z) = ax + by + cz$. Dados os vetores $u = (1, 2, 3)$, $v = (-1, 2, 3)$ e $w = (1, -2, 3)$, determine a, b e c de tal modo que se tenha $f(u) = 1$, $f(v) = 0$ e $f(w) = 0$.

4.8. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador linear definido por $A(x, y) = (5x + 4y, -3x - 2y)$. Ache vetores não-nulos $u = (x, y)$ e $v = (s, t)$ tais que $Au = u$ e $Av = 2v$. São únicas as soluções? Será possível achar $w \neq 0$ em $\mathbb{R}^2$ com $Aw = \alpha w$, onde $\alpha \neq 1$ e $\alpha \neq 2$?

4.9. Dê as expressões dos funcionais lineares $f, g, h\colon \mathbb{R}^3 \to \mathbb{R}$ que formam a base dual em $(\mathbb{R}^3)^*$ da base $\{u, v, w\} \subset \mathbb{R}^3$, onde $u = (1, 1, 1)$, $v = (1, -1, 1)$ e $w = (1, 1 - 1)$. (Vide Exercício 4.20.)

4.10. Tem-se uma transformação linear $A\colon \mathbb{R}^2 \to \mathbb{R}^3$. Sabe-se que $A(-1, 1) = (1, 2, 3)$ e $A(2, 3) = (1, 1, 1)$. Pede-se a matriz $\mathbf{a} \in M(3 \times 2)$ de A relativamente às bases canônicas de $\mathbb{R}^2$ e $\mathbb{R}^3$.

4.11. Prove que uma transformação linear $A\colon E \to F$ transforma todo conjunto convexo $C \subset E$ num conjunto convexo $A(C) \subset F$.

4.12. Determine a expressão do operador linear $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, sabendo que, para todo $v = (x, y)$, o segmento de reta que liga v a $Av = (x', y')$ é horizontal e tem seu ponto médio sobre a reta $y = x$. Qual é a imagem do eixo vertical pelo operador A?

4.13. Prove que os operadores lineares $E_{11}, E_{12}, E_{21}, E_{22}\colon \mathbb{R}^2 \to \mathbb{R}^2$, definidos por $E_{11}(x, y) = (x, 0)$, $E_{12}(x, y) = (0, x)$, $E_{21}(x, y) = (y, 0)$,

$E_{22}(x, y) = (0, y)$, constituem uma base do espaço vetorial $\mathcal{L}(\mathbb{R}^2)$. Prove ainda que outra base deste espaço pode ser formada com os operadores A, B, C, I, onde $A(x, y) = (x + 3y, y)$, $B(x, y) = (x, 0)$, $C(x, y) = (x + y, x - y)$ e $I(x, y) = (x, y)$.

4.14. Verifique que as funções definidas nos Exercícios 3.32 e 3.33 são funcionais lineares.

4.15. Seja $A \colon E \to F$ uma transformação linear
(a) Se os vetores $Av_1, \ldots, Av_m \in F$ são L.I., prove que $v_1, \ldots, v_m \in E$ também são L.I..
(b) Se $F = E$ e os vetores $Av_1, \ldots, Av_m$ geram E, prove que $v_1, \ldots, v_m$ geram E.
(c) Valem as recíprocas de (a) e (b)? Seria (b) verdadeira com $F \neq E$?

4.16. Quais das transformações abaixo são lineares?

(a) $A \colon \mathbb{R}^3 \to \mathbb{R}^3$, $A(x, y, z) = (x, 2^y, 2^z)$.

(b) $A \colon \mathbb{R}^3 \to \mathbb{R}^3$, $A(x, y, z) = (3x, a, 5z)$, onde $a \in \mathbb{R}$.

(c) $A \colon \mathbb{R}^4 \to \mathbb{R}^3$, $A(x, y, z, w) = (x - w, y - w, x + z)$.

(d) $A \colon M(n \times n) \to \mathbb{R}^n$, $A([a_{ij}]) = (a_{11}, a_{22}, \ldots, a_{nn})$.

(e) $A \colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$, $Af = 3f'' - 2f' + 1$.

(f) $A \colon M(2 \times 2) \to \mathbb{R}$, $A\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$.

4.17. Sejam $A \colon E \to F$ uma transformação linear e $E' \subset E$, $F' \subset F$ subespaços vetoriais. Prove que $A(E') = \{Av; v \in E'\}$ é um subespaço de F e $A^{-1}(F') = \{v \in E; Av \in F'\}$ é um subespaço de E. Se $V \subset E$ e $W \subset F$ são variedades afins, prove que os conjuntos $A(V) \subset F$ e $A^{-1}(W) \subset E$, definidos analogamente, são também variedades afins.

4.18. No exercício anterior, prove que se E' tem dimensão finita então $\dim A(E')$ é finita e $\dim A(E') \leq \dim E'$. Dê um exemplo de um operador não identicamente nulo $A \colon \mathbb{R}^2 \to \mathbb{R}^2$ e um subespaço $E' \subset \mathbb{R}^2$ tal que $\dim A(E') < \dim E'$. Prove também que se E e F' têm dimensão finita e A é sobrejetiva então $\dim A^{-1}(F') \geq \dim F'$. Dê um exemplo em que $A \neq 0$ e $\dim A^{-1}(F') > \dim F'$. Dê também um exemplo (com $\dim E = \infty$), onde $\dim F'$ é finita mas $\dim A^{-1}(F') = \infty$.

4.19. Dados os espaços vetoriais E, F, prove que $\mathcal{L}(E; F)$ é um subespaço vetorial de $\mathcal{F}(E; F)$. (Vide Exercício 1.15.)

4.20. Seja $\mathcal{V} = \{v_1, \ldots, v_n\}$ uma base do espaço vetorial E. Para cada $i = 1, 2, \ldots, n$, seja $f_i \colon E \to \mathbb{R}$ o funcional linear determinado (conforme o Teorema 4.1) pelas condições $f_i(v_i) = 1$, $f_i(v_j) = 0$ se $j \neq i$. Prove que $\{f_1, \ldots, f_n\}$ é uma base de $E^* = \mathcal{L}(E; \mathbb{R})$ (chamada a *base dual* da base $\mathcal{V}$). Mostre que se tem $f_i(v) = x_i$ para todo $v = x_1 v_1 + \cdots + x_n v_n \in E$.

4.21. Seja $f \colon \mathbb{R}^2 \to \mathbb{R}$ um funcional linear. Sabendo que $f(1,1) = 3$ e $f(2,3) = 1$, calcule $f(1,0)$ e $f(0,1)$.

4.22. Seja $A \colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador linear dado por $A(x,y) = (ax + by, cx + dy)$, com $ad - bc \neq 0$. Prove:

(1) Para todo $v \neq 0$ em $\mathbb{R}^2$, tem-se $A.v \neq 0$.

(2) Toda reta $R \subset \mathbb{R}^2$ (variedade afim de dimensão 1) é transformada por A numa reta.

(3) A transforma retas paralelas em retas paralelas.

4.23. Determine α de modo que as retas perpendiculares em $\mathbb{R}^2$, de equações $y = \alpha x$ e $y = -x/\alpha$ sejam transformadas em retas perpendiculares pelo operador linear $A \colon \mathbb{R}^2 \to \mathbb{R}^2$, dado por $A(x,y) = (2x + 3y, x - 2y)$.

4.24. Sejam E, F espaços vetoriais de dimensão finita. Dados os vetores $v_1, \ldots, v_m \in E$ e $w_1, \ldots, w_m \in F$, a fim de que exista uma transformação linear $A \colon E \to F$ com $Av_1 = w_1, \ldots, Av_m = w_m$, é necessário e suficiente que, para toda combinação linear nula $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$, se tenha também $\alpha_1 w_1 + \cdots + \alpha_m w_m = 0$.

4.25. Seja v um vetor não-nulo de um espaço vetorial E, de dimensão finita. Dado qualquer espaço vetorial $F \neq \{0\}$, mostre que existe uma transformação linear $A \colon E \to F$ tal que $Av \neq 0$.

4.26. Seja E um espaço vetorial de dimensão finita. Dada uma base $\mathcal{F} = \{f_1, \ldots, f_n\} \subset E^*$, mostre que existe uma base $\{v_1, \ldots, v_n\} \subset E$ da qual $\mathcal{F}$ é dual. (Veja Exercício 4.20.)

4.27. Seja Y um conjunto de geradores do espaço vetorial E. Se as transformações lineares $A, B \colon E \to F$ são tais que $Aw = Bw$ para todo $w \in Y$, prove que $Av = Bv$ para todo $v \in E$.

4.28. Seja $X = \{v_1, \dots, v_m\}$ um conjunto L.I. no espaço vetorial E, de dimensão finita. Dados arbitrariamente os vetores $w_1, \dots, w_m$ no espaço vetorial F, prove que existe uma transformação linear $A \colon E \to F$ tal que $Av_1 = w_1, \dots, Av_m = w_m$. A é única se, e somente se, X é uma base de E.

4.29. Uma transformação $T \colon E \to F$, entre espaços vetoriais, chama-se *afim* quando se tem $T((1-t)u+tv) = (1-t)Tu+tTv$ para quaisquer $u, v \in E$ e $t \in \mathbb{R}$. Dada a transformação afim $T \colon E \to F$, prove:

(a) Toda variedade afim $V \subset E$ é transformada por T numa variedade afim $V' \subset F$.

(b) Se $T \cdot 0 = 0$, então, escrevendo $\alpha v = (1-\alpha)0 + \alpha v$, resulta que $T(\alpha v) = \alpha \cdot Tv$ para quaisquer $\alpha \in \mathbb{R}$, $v \in E$.

(c) Supondo ainda $T \cdot 0 = 0$, a relação $T\left(\frac{1}{2}(u+v)\right) = \frac{1}{2}(Tu + Tv)$, implica que $T(u+v) = Tu + Tv$ para quaisquer $u, v \in E$.

(d) Para todo $b \in F$, a transformação $S \colon E \to F$, definida por $Sv = Tv + b$ também é afim.

Conclua que $T \colon E \to F$ é uma transformação afim se, e somente se, existem $A \in \mathcal{L}(E; F)$ e $b \in F$ tais que $Tv = Av + b$ para todo $v \in E$.

4.30. Seja $H \subset \mathbb{R}^n$ um subespaço vetorial de dimensão $n-1$. Tome uma base $\mathcal{V} \subset H$, formada pelos vetores $v_1, \dots, v_{n-1}$, onde $v_i = (\alpha_i, \dots, \alpha_{in})$, $i = 1, \dots, n-1$. Use o fato de que o sistema de equações lineares $\sum_{j=1}^{n} \alpha_{ij} x_j = 0$, com $n-1$ equações e n incógnitas, admite uma solução não-trivial para concluir que existem n números $a_1, \dots, a_n$ tais que $v = (x_1, \dots, x_n)$ pertence a H se, e somente se, $a_1 x_1 + \cdots + a_n x_n = 0$. (Todo subespaço vetorial de $\mathbb{R}^n$ com dimensão $n-1$ é um hiperplano, isto é, é o conjunto das soluções de uma equação linear homogênea.)

5

Produto de Transformações Lineares

O produto de transformações lineares, que introduziremos nesta seção, é um exemplo concreto de estrutura algébrica que apresenta variados e interessantes fenômenos, não encontrados nas operações entre números ou entre vetores.

Dadas as transformações lineares $A\colon E \to F$, $B\colon F \to G$, onde o domínio de B coincide com o contra-domínio de A, define-se o *produto* $BA\colon E \to G$ pondo, para cada $v \in E$, $(BA)v = B(Av)$,

$$E \xrightarrow{A} F \xrightarrow{B} G, \qquad E \xrightarrow{BA} G$$

Vê-se imediatamente que BA é uma transformação linear. Observe-se também que BA nada mais é do que a composta $B \circ A$ das funções B e A. Segue-se então dos princípios gerais que se $C\colon G \to H$ é outra transformação linear, vale a

Associatividade: $(CB)A = C(BA)$.

A linearidade tampouco é necessária para mostrar que, dadas $A\colon E \to F$ e $B, C\colon F \to G$, tem-se a

Distributividade à esquerda: $(B + C)A = BA + CA$, que decorre simplesmente da definição de $B + C$.

Usando a linearidade de $C\colon F \to G$, vê-se que, dadas $A, B\colon E \to F$, vale a

Distributividade à direita: $C(A + B) = CA + CB$.

Com efeito, para todo $v \in E$, tem-se

$$
\begin{aligned}
[C(A + B)]v &= C[(A + B)v] = C(Av + Bv) = C(Av) + C(Bv) \\
&= (CA)v + (CB)v = (CA + CB)v.
\end{aligned}
$$

Exemplo 5.1. Sejam $f, g, h \colon \mathbb{R} \to \mathbb{R}$ definidas por $f(x) = x$, $g(x) = x + 1$ e $h(x) = x^2$. Então $[h \circ (f + g)](x) = 4x^2 + 4x + 1$, enquanto $[(h \circ f) + (h \circ g)](x) = 2x^2 + 2x + 1$, logo $h \circ (f + g) \neq h \circ f + h \circ g$. Isto se dá porque h não é linear.

Outra conseqüência da linearidade de B é a

Homogeneidade: $B(\alpha A) = \alpha(BA)$, válida para $\alpha \in \mathbb{R}$, $A \colon E \to F$ e $B \colon F \to G$ quaisquer.

Evidentemente, dada $A \colon E \to F$, tem-se $AI_E = A = I_F A$, de modo que as aplicações identidade $I_E \colon E \to E$, $I_F \colon F \to F$ são elementos neutros para a multiplicação, cada uma delas do lado apropriado.

Diferenças notáveis entre o produto de transformações lineares e o produto de números reais são as ausências da comutatividade, da lei do corte e da inversa multiplicativa para uma transformação $\neq 0$, além da presença de transformações nilpotentes, para as quais tem-se $A^n = 0$ com $A \neq 0$. Deve-se ainda mencionar a restrição de que o produto BA só está definido quando A toma valores no domínio de B. Esta restrição desaparece, naturalmente, quando se trata de operadores lineares no mesmo espaço E: então o produto BA está definido quaisquer que sejam $A, B \in \mathcal{L}(E)$.

Exemplo 5.2. Sejam $P, R \colon \mathbb{R}^2 \to \mathbb{R}^2$ respectivamente a projeção ortogonal sobre a reta $y = x$ e a rotação de um ângulo de $90°$ em torno da origem. Então, para todo $v = (x, y) \in \mathbb{R}^2$, tem-se $Pv = \frac{1}{2}(x + y, x + y)$, $Rv = (-y, x)$. Segue-se que

$$
RPv = \frac{1}{2}(-x - y, x + y)
$$

e

$$
PRv = \frac{1}{2}(x - y, x - y).
$$

Portanto $RPv \neq PRv$ para todo v, exceto para $v = (0, 0)$. Observe que bastaria que $RPv \neq PRv$ para um *único* v a fim de termos $RP \neq PR$.

Exemplo 5.3. Seja $P\colon \mathbb{R}^2 \to \mathbb{R}^2$ a projeção ortogonal sobre uma certa reta r. Para todo v sobre a reta r, tem-se $Pv = v$. Assim, para qualquer $v \in \mathbb{R}^2$, tem-se $PPv = Pv$, pois Pv está sobre r. Noutras palavras, vale $PP = P$, ou seja $PP = PI$, embora $P \neq I$. Assim, não é permitido cortar o fator P à esquerda em ambos os membros da igualdade $PP = PI$. Segue-se que não existe $Q \in \mathcal{L}(\mathbb{R}^2)$ tal que $QP = I$. Com efeito, se um tal operador Q existisse, de $PP = P$ concluiríamos $QPP = QP$, isto é, $IP = I$, donde $P = I$.

Exemplo 5.4. Sejam $P, Q\colon \mathbb{R}^2 \to \mathbb{R}^2$ projeções ortogonais sobre duas retas do plano, uma das quais é perpendicular à outra. Todo vetor $v \in \mathbb{R}^2$ é a diagonal de um retângulo que tem Pv e Qv como lados. (Veja Fig. 5.1.)

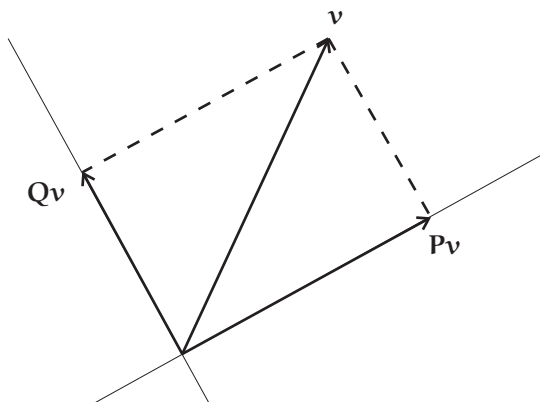



Figura 5.1.

Segue-se então que $v = Pv + Qv$ para todo $v \in \mathbb{R}^2$, ou seja, $P + Q = I$ e $Q = I - P$. Portanto $PQ = P(I - P) = P - P^2 = P - P = 0$. Obtemos assim dois operadores não-nulos P, Q com $PQ = 0$. É possível mesmo que um operador não-nulo $A \in \mathcal{L}(\mathbb{R}^2)$ cumpra $A^2 = 0$. Basta pôr $A(x, y) = (x - y, x - y)$.

Um operador A chama-se *nilpotente* quando, para algum $n \in \mathbb{N}$, tem-se $A^n = 0$. Um exemplo significativo de operador nilpotente é a derivação $D\colon \mathcal{P}_n \to \mathcal{P}_n$. Para todo polinômio p de grau $\leq n$ tem-se $D^{n+1}p = 0$, logo $D^{n+1} = 0$.

Exemplo 5.5. Se $R_\alpha, R_\beta\colon \mathbb{R}^2 \to \mathbb{R}^2$ são rotações em torno da origem com ângulos α e β respectivamente, então $R_\alpha \cdot R_\beta = R_{\alpha+\beta}$. (Isto pode

ser visto geometricamente na Fig. 5.2 ou usando as fórmulas de $\cos(\alpha+\beta)$ e $\operatorname{sen}(\alpha+\beta)$). Se $S\colon \mathbb{R}^2 \to \mathbb{R}^2$ é a reflexão em torno de uma reta então $S \cdot S = I$. Isto se segue da expressão $S = 2P - I$, levando em conta que $P \cdot P = P$, mas também pode ser visto geometricamente (com mais facilidade).

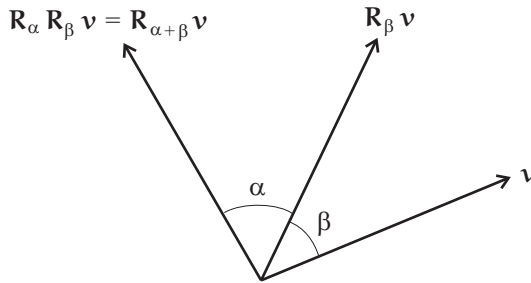



Figura 5.2 – Rotações do plano.

Exercícios

5.1. Verifique explicitamente que o produto BA de transformações lineares é ainda uma transformação linear.

5.2. Considere os operadores lineares $R, S, P\colon \mathbb{R}^2 \to \mathbb{R}^2$, onde R é a rotação de 30° em torno da origem, S é a reflexão em torno da reta $y = 2x$ e P é a projeção ortogonal sobre a mesma reta.

(i) Mostre que se tem $PS = SP = P$.

(ii) Verifique a igualdade $RSR = S$.

(iii) Mostre que R não comuta com S nem com P.

(iv) Determine todos os vetores v tais que $PRv = 0$ e $RPv \neq 0$.

5.3. Dado o operador linear $A\colon E \to E$, seja N o conjunto dos vetores $v \in E$ tais que $Av = 0$. Mostre que N é um subespaço vetorial de E. Prove que $A^2 = 0$ se, e somente se, para todo $v \in E$ tem-se $Av \in N$.

5.4. Sejam $R, R': \mathbb{R}^2 \to \mathbb{R}^2$ respectivamente as rotações de ângulos θ e θ' em torno da origem. Partindo do fato de que o produto RR' é a rotação de ângulo $\theta + \theta'$, use o Exemplo 4.2 para obter as fórmulas clássicas $\cos(\theta + \theta') = \cos\theta \cdot \cos\theta' - \operatorname{sen}\theta \cdot \operatorname{sen}\theta'$ e $\operatorname{sen}(\theta + \theta') = \operatorname{sen}\theta \cdot \cos\theta' + \operatorname{sen}\theta' \cdot \cos\theta$.

5.5. Seja $A: E \to E$ um operador nilpotente. Prove que existe algum vetor $v \neq 0$ em E tal que $Av = 0$.

5.6. Dados os operadores $A, B: \mathbb{R}^2 \to \mathbb{R}^2$ dados por $A(x, y) = (x+y, 0)$ e $B(x, y) = (-y, x)$, obtenha as expressões dos operadores $A + B$, AB, BA, A^2 e B^2. Descreva geometricamente esses cinco operadores. (Exemplo: A é a projeção sobre o eixo x paralelamente a uma certa reta. (Qual?))

5.7. Seja $A: \mathbb{R}^3 \to \mathbb{R}^3$ dado por $A(x, y, z) = (ay + bz, cz, 0)$. Mostre que $A^3 = 0$.

5.8. Sejam $A, B, C, D: \mathbb{R}^2 \to \mathbb{R}^2$ os operadores dados por $A(x, y) = (x, 0)$, $B(x, y) = (-y, x)$, $C(x, y) = (0, y)$ e $D(x, y) = (y, -x)$. Determine o operador $ABCD$.

5.9. Considere as transformações lineares $A: \mathbb{R}^2 \to \mathbb{R}^3$ e $B: \mathbb{R}^3 \to \mathbb{R}^2$, definidas por: $A(x, y) = (x, y, x+y)$ e $B(x, y, z) = (ax + (a-1)y + (1-a)z, -bx + (1-b)y + bz)$. Determine o operador $BA: \mathbb{R}^2 \to \mathbb{R}^2$.

5.10. Dado o operador $A: \mathbb{R}^2 \to \mathbb{R}^2$, com $A(x, y) = (3x - 2y, 2x + 7y)$, ache um vetor não-nulo $v = (x, y)$ tal que $Av = 5v$.

5.11. Sejam $A, B: E \to E$ operadores lineares. Suponha que existam vetores $u, v \in E$ tais que Au e Av sejam L.D.. Prove que BAu e BAv são L.D.. Se a dimensão de E for igual a 2, prove também que ABu e ABv são L.D.. (Sugestão: se u e v são L.D. o fato é óbvio. Caso contrário, u e v formam uma base de E. Exprima Bu e Bv em termos dessa base e depois aplique A.)

5.12. Sejam $A, B: \mathbb{R}^3 \to \mathbb{R}^3$ definidos por $A(x, y, z) = (x, y, 0)$ e $B(x, y, z) = (x + z, y, 0)$. Obtenha vetores $u, v \in \mathbb{R}^3$ tais que Au e Av sejam L.D. porém ABu e ABv sejam L.I..

5.13. No espaço vetorial $\mathcal{P}$ dos polinômios, considere os operadores lineares $D, A: \mathcal{P} \to \mathcal{P}$ de derivação $(Dp(x) = p'(x))$ e multiplicação por x $(Ap(x) = xp(x))$ respectivamente. Determine $DA - AD$.

5.14. Seja $A: \mathbb{R}^2 \to \mathbb{R}^2$ o operador linear definido por $A(x, y) = (ax + by, cx + dy)$. Verifique que $A^2 - (a + d)A = (bc - ad)I$. Use esta igualdade para provar que, se $ad - bc \neq 0$, existe um operador linear $B: \mathbb{R}^2 \to \mathbb{R}^2$ tal que $BA = AB = I$.

5.15. Seja $C(A)$ o conjunto dos operadores lineares $X: E \to E$ que comutam com o operador $A \in \mathcal{L}(E)$ (isto é, $AX = XA$). Prove que $C(A)$ é um subespaço vetorial de $\mathcal{L}(E)$ e que $X, Y \in C(A) \Rightarrow XY \in C(A)$.

5.16. Seja $A: \mathbb{R}^2 \to \mathbb{R}^2$ um operador linear que não é da forma αI. Prove:
(a) Existe $w \in \mathbb{R}^2$ tal que $\{w, Aw - w\} \subset \mathbb{R}^2$ é uma base.
(b) Se $P: \mathbb{R}^2 \to \mathbb{R}^2$ é o operador que a cada $v = xw + y(Aw - w) \in \mathbb{R}^2$ faz corresponder $Pv = xw$ então $AP \neq PA$.

Conclua que os únicos operadores lineares em $\mathbb{R}^2$ que comutam com todos os demais são os da forma αI.

5.17. Estenda o exercício anterior para todos os espaços vetoriais de dimensão finita (maior do que 1) em vez de $\mathbb{R}^2$.

5.18. Sejam $A: E \to F$ uma transformação linear e $B: F \to E$ uma função tal que $AB = I_F$ e $BA = I_E$. Prove que B é uma transformação linear.

5.19. Seja E um espaço vetorial de dimensão n. Para todo $k = 2, \ldots, n$, exiba um operador linear $A: E \to E$ tal que $A^k = 0$ mas $A^j \neq 0$ se $j < k$.

5.20. Sejam $A, P: E \to E$ operadores lineares não-nulos tais que $AP = 0$. Prove que existem vetores diferentes de zero $u \neq v$ com $Au = Av$.

5.21. Sejam $A: \mathbb{R}^2 \to \mathbb{R}^2$ um operador linear e $P: \mathbb{R}^2 \to \mathbb{R}^2$ a projeção ortogonal sobre uma reta r (passando pela origem). Prove que as seguintes afirmações são equivalentes:
(a) Para todo $v \in r$ tem-se $Av \in r$.
(b) $PAP = AP$.

5.22. Seja $X: F \to G$ uma transformação linear tal que $Xw \neq 0$ para todo $w \neq 0$ em F. Prove que se $A, B \in \mathcal{L}(E; F)$ cumprem $XA = XB$ então $A = B$.

5.23. Seja $X: E \to F$ uma transformação linear com a seguinte propriedade: para cada $w \in F$ existe (pelo menos) um vetor $v \in E$ tal que $Xv = w$. Suponha que $A, B \in \mathcal{L}(F; G)$ cumprem a igualdade $AX = BX$. Prove que $A = B$.

5.24. Sejam $A, B: E \to F$ transformações lineares tais que, para todo operador linear $X: F \to F$, tem-se $XA=XB$. Prove que $A=B$.

6

Núcleo e Imagem

Nesta seção, será examinada com cuidado a possibilidade de uma transformação linear admitir ou não uma inversa. Veremos que isto está associado à existência e à unicidade da solução de um sistema de equações lineares. Será introduzido o conceito de isomorfismo, que dará um sentido preciso à afirmação de que dois espaços vetoriais de mesma dimensão são algebricamente indistinguíveis. Tudo começa com o núcleo e a imagem de uma transformação.

A toda transformação linear $A\colon E \to F$ estão associados dois subespaços vetoriais indispensáveis para estudar o comportamento de A: o núcleo de A, que é um subespaço de E, e a imagem de A, que é um subespaço de F.

A *imagem de* A é o subconjunto $\mathcal{I}m(A) \subset F$, formado por todos os vetores $w = Av \in F$ que são imagens de elementos de E pela transformação A.

A noção de imagem tem sentido seja qual for a função $A\colon E \to F$, seja linear ou não. Quando A é linear, então $\mathcal{I}m(A)$ é um subespaço vetorial de F, como se vê facilmente.

Se $\mathcal{I}m(A) = F$, diz-se que a transformação A é *sobrejetiva*. Isto significa que, para qualquer $w \in F$ dado, pode-se achar $v \in E$ tal que $A \cdot v = w$.

Seja $X \subset E$ um conjunto de geradores do espaço vetorial E. A imagem da transformação linear $A\colon E \to F$ é o subespaço vetorial de F gerado pelos vetores Av, $v \in X$. Em particular, A é sobrejetiva se, e somente se, transforma X num conjunto de geradores de F.

Se $v_1, \ldots, v_n$ geram E os vetores $Av_1, \ldots, Av_n$ geram $\mathcal{I}m(A)$. Segue-se que a dimensão de $\mathcal{I}m(A)$ é menor do que ou igual à dimensão do domínio de A. Este fato será tornado mais preciso a seguir. (V. Teorema 6.6.)

Exemplo 6.1. Dado um sistema linear de m equações a n incógnitas

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m\,, \end{aligned}$$

seja $A\colon \mathbb{R}^n \to \mathbb{R}^m$ a transformação linear cuja matriz nas bases canônicas de $\mathbb{R}^n$ e $\mathbb{R}^m$ é $\mathbf{a} = [a_{ij}]$. Isto significa, como sabemos, que, para $j = 1, \ldots, n$, os vetores

$$v_j = Ae_j = \sum_{i=1}^{m} a_{ij}e_i = (a_{1j}, \ldots, a_{mj}) \in \mathbb{R}^m$$

são os vetores-coluna da matriz $\mathbf{a}$. Em termos da transformação linear A, o sistema acima pode ser interpretado como o problema de achar um vetor $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ tal que $Ax = b$, onde $b = (b_1, \ldots, b_m)$. Portanto o sistema admite solução se, e somente se, o vetor b pertence à imagem da transformação linear A, o que equivale a dizer que os conjuntos $\{v_1, \ldots, v_n\}$ e $\{v_1, \ldots, v_n, b\}$ geram ambos o mesmo subespaço $\mathcal{I}m(A)$.

Exemplo 6.2. Um funcional linear $f\colon E \to \mathbb{R}$ é sobrejetivo ou é igual a zero, pois $\{0\}$ e $\mathbb{R}$ são os únicos subespaços vetoriais de $\mathbb{R}$. A derivação $D\colon C^k(\mathbb{R}) \to C^{k-1}(\mathbb{R})$ é sobrejetiva, e o mesmo se dá com o operador $D\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ e com a transformação linear $D\colon \mathcal{P}_n \to \mathcal{P}_{n-1}$. Se $P\colon \mathbb{R}^2 \to \mathbb{R}^2$ é a projeção ortogonal sobre uma reta r, a imagem de P é essa reta r.

Uma transformação linear $B\colon F \to E$ chama-se uma *inversa à direita* da transformação linear $A\colon E \to F$ quando se tem $AB = I_F$, ou seja, quando $A(Bw) = w$ para todo $w \in F$.

Teorema 6.1. *A fim de que uma transformação linear* $A\colon E \to F$, *entre espaços vetoriais de dimensão finita, possua uma inversa à direita* $B \in \mathcal{L}(F; E)$ *é necessário e suficiente que* A *seja sobrejetiva.*

Demonstração: Se A admite uma inversa à direita $B\colon F \to E$ então, para todo $w \in F$ tem-se $A(Bw) = w$, logo $w = A \cdot v$, onde $v = Bw$, e A é sobrejetiva. (Aqui não se usou a linearidade de A, e muito menos a finitude das dimensões de E e F.) Suponhamos, em seguida, que A seja sobrejetiva. A fim de definir uma transformação linear $B\colon F \to E$ com $A(Bw) = w$ para todo $w \in F$, tomamos uma base $\mathcal{B} = \{w_1, \ldots, w_m\} \subset F$. Como A é sobrejetiva, podemos escolher vetores $v_1, \ldots, v_m \in E$ tais que $Av_1 = w_1, \ldots, Av_m = w_m$. Pelo Teorema 4.1, existe uma transformação linear $B\colon F \to E$ tal que $Bw_1 = v_1, \ldots, Bw_m = v_m$. Afirmamos que, para todo $w \in F$, tem-se $A(Bw) = w$. Com efeito, sendo $\mathcal{B}$ uma base, podemos escrever $w = \beta_1 w_1 + \cdots + \beta_m w_m$, portanto

$$\begin{aligned} A(Bw) &= A(\beta_1 Bw_1 + \cdots + \beta_m Bw_m) \\ &= A(\beta_1 v_1 + \cdots + \beta_m v_m) \\ &= \beta_1 Av_1 + \cdots + \beta_m Av_m \\ &= \beta_1 w_1 + \cdots + \beta_m w_m = w. \end{aligned}$$

□

Exemplo 6.3. Uma transformação linear sobrejetiva $A\colon E \to F$ pode admitir mais de uma inversa à direita $B\colon F \to E$. Um exemplo simples é dado pela transformação linear $A\colon \mathbb{R}^3 \to \mathbb{R}^2$, definida por $A(x,y,z) = (x,y)$. Fixados arbitrariamente $a, b \in \mathbb{R}$, a transformação linear $B\colon \mathbb{R}^2 \to \mathbb{R}^3$, definida por $B(x,y) = (x, y, ax + by)$, é uma inversa à direita para A. Variando os números a e b, obtemos infinitas possibilidades para B.

Exemplo 6.4. Uma inversa à direita para a derivação $D\colon \mathcal{P}_{n+1} \to \mathcal{P}_n$ é a transformação linear $J\colon \mathcal{P}_n \to \mathcal{P}_{n+1}$, que a cada polinômio $p(x) = a_o + a_1 x + \cdots + a_n x^n$ de grau $\leq n$ faz corresponder o polinômio

$$Jp(x) = a_o x + \frac{a_1}{2} x^2 + \cdots + \frac{a_n}{n+1} x^{n+1}.$$

O *núcleo* da transformação linear $A\colon E \to F$ é o conjunto dos vetores $v \in E$ tais que $Av = 0$. Usaremos a notação $\mathcal{N}(A)$ para representar o núcleo de A. É fácil ver que $\mathcal{N}(A)$ é um subespaço vetorial de E.

Uma transformação linear $A\colon E \to F$ chama-se *injetiva* quando $v \neq v'$ em $E \Rightarrow Av \neq Av'$ em F. Equivalentemente: $Av = Av' \Rightarrow v = v'$. Esta noção tem sentido para qualquer função $A\colon E \to F$, seja ela

linear ou não. No caso linear, porém, o teorema abaixo simplifica a verificação da injetividade.

Teorema 6.2. *A fim de que uma transformação linear* $A\colon E \to F$ *seja injetiva é necessário e suficiente que seu núcleo* $\mathcal{N}(A)$ *contenha apenas o vetor nulo.*

Demonstração: Seja A injetiva. Então $v \in \mathcal{N}(A) \Rightarrow A \cdot v = 0 = A \cdot 0 \Rightarrow v = 0$, logo $\mathcal{N}(A) = \{0\}$. Reciprocamente, seja $\mathcal{N}(A) = \{0\}$. Então $Av = Av' \Rightarrow A(v - v') = Av - Av' = 0 \Rightarrow v - v' \in \mathcal{N}(A) \Rightarrow v - v' = 0 \Rightarrow v = v'$, logo A é injetiva. □

Teorema 6.3. *Uma transformação linear é injetiva se, e somente se, leva vetores* L.I. *em vetores* L.I..

Demonstração: Seja $A\colon E \to F$ uma transformação linear injetiva. Se os vetores $v_1, \ldots, v_n \in E$ são linearmente independentes, vamos provar que suas imagens $Av_1, \ldots, Av_n$ são vetores linearmente independentes em F. Com efeito, se $\alpha_1 \cdot Av_1 + \cdots + \alpha_n \cdot Av_n = 0$ então $A(\alpha_1 v_1 + \cdots + \alpha_n v_n) = 0$, logo $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ pois A é injetiva. Como $v_1, \ldots, v_n$ são L.I., segue-se que $\alpha_1 = \cdots = \alpha_n = 0$, portanto $Av_1, \ldots, Av_n$ são L.I.. Reciprocamente se a transformação linear $A\colon E \to F$ leva vetores L.I. em vetores L.I. então $v \neq 0$ em $E \Rightarrow \{v\}$ L.I. $\Rightarrow \{Av\}$ L.I. $\Rightarrow Av \neq 0$, portanto $\mathcal{N}(A) = \{0\}$ e A é injetiva. □

Segue-se deste teorema que se E tem dimensão finita n e $A\colon E \to F$ é uma transformação linear injetiva então $\dim F \geq n$. Assim, por exemplo, não existe uma transformação linear injetiva de $\mathbb{R}^3$ em $\mathbb{R}^2$.

Teorema 6.4. *Seja* $A\colon E \to F$ *uma transformação linear. Para todo* $b \in \mathcal{I}m(A)$, *o conjunto* $V = \{x \in E; Ax = b\}$, *formado pelas soluções do sistema linear* $Ax = b$, *é uma variedade afim em* E, *paralela ao núcleo* $\mathcal{N}(A)$.

Demonstração: Fixemos $x_0 \in V$, isto é, com $Ax_0 = b$. Afirmamos que $V = x_0 + \mathcal{N}(A)$. Com efeito, $v \in \mathcal{N}(A) \Rightarrow A(x_0 + v) = Ax_0 + Av =$

$b + 0 = b \Rightarrow x_0 + v \in V$. Logo $x_0 + \mathcal{N}(A) \subset V$. Reciprocamente,

$$\begin{aligned} x \in V \Rightarrow x &= x_0 + (x - x_0) = x_0 + v \\ &\Rightarrow b = Ax = A(x_0 + v) = Ax_0 + Av = b + Av \\ &\Rightarrow b = b + Av \\ &\Rightarrow Av = 0 \\ &\Rightarrow x = x_0 + v \in x_0 + \mathcal{N}(A). \end{aligned}$$

Logo $V \subset x_0 + \mathcal{N}(A)$. $\square$

Observação. Geometricamente, o Teorema 6.4 significa que o espaço vetorial E se exprime como uma reunião de lâminas paralelas $V = x_0 + \mathcal{N}(A)$, cada uma das quais é uma variedade afim que se transforma por A num único ponto $b \in \mathcal{I}m(A)$. Este ponto, naturalmente, varia quando se passa de uma lâmina para outra. (Veja Fig. 6.1.)

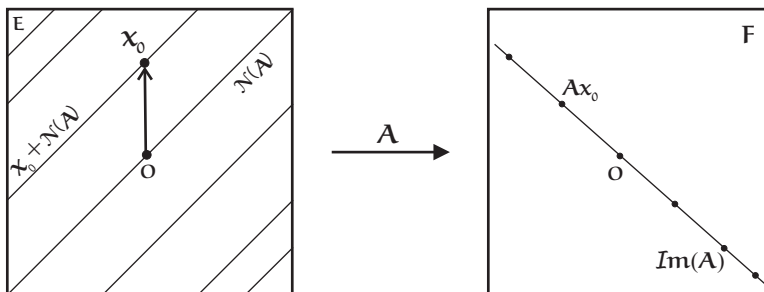



Figura 6.1.

Algebricamente, o Teorema 6.4 significa que, para cada $b \in \mathcal{I}m(A)$, obtêm-se todas as soluções $x \in E$ do sistema linear $Ax = b$ assim: acha-se uma "solução particular" x_0 desse sistema e a solução geral $x = x_o + v$ é a soma dessa solução particular com a "solução geral v do sistema homogêneo associado" $Ax = 0$. Naturalmente, esta última é um elemento qualquer do núcleo de A. Se $b \notin \mathcal{I}m(A)$ então o sistema $Ax = b$, evidentemente, não possui solução.

Exemplo 6.5. O núcleo de uma rotação ou de uma reflexão no plano $\mathbb{R}^2$ reduz-se a $\{0\}$. O núcleo da projeção ortogonal $P \colon \mathbb{R}^2 \to \mathbb{R}^2$ sobre a reta r é a reta que contém 0 e é perpendicular a r. O núcleo da

derivação $D\colon C^k(\mathbb{R}) \to C^{k-1}(\mathbb{R})$ é o subespaço uni-dimensional de $C^k(\mathbb{R})$ formado pelas funções constantes. O núcleo de um funcional linear não-nulo $\varphi\colon E \to \mathbb{R}$ é um hiperplano $H \subset E$.

Sejam $A\colon E \to F$ e $B\colon F \to E$ transformações lineares. Diz-se que B é uma *inversa à esquerda* de A quando $BA = I_E$, isto é, quando $B(Av) = v$ para todo $v \in E$.

Exemplo 6.6. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^3$ definida por $A(x, y) = (x + 2y, 2x + 3y, 3x + 4y)$. A transformação linear $B\colon \mathbb{R}^3 \to \mathbb{R}^2$, dada por

$$B(x, y, z) = (-3x + 2y, 2x - y)$$

cumpre a relação

$$\begin{aligned} B(A(x, y)) &= B(x + 2y, 2x + 3y, 3x + 4y) \\ &= (-3(x + 2y) + 2(2x + 3y), 2(x + 2y) - (2x + 3y)) \\ &= (x, y) \end{aligned}$$

para qualquer $(x, y) \in \mathbb{R}^2$. Logo B é uma inversa à esquerda para A.

Exemplo 6.7. Uma transformação linear pode admitir uma infinidade de inversas à esquerda. Por exemplo, seja $A\colon \mathbb{R}^2 \to \mathbb{R}^3$ dada por $A(x, y) = (x, y, 0)$. Para quaisquer $a, b \in \mathbb{R}$, a transformação linear $B\colon \mathbb{R}^3 \to \mathbb{R}^2$, dada por $B(x, y, z) = (x + az, y + bz)$ é uma inversa à esquerda de A, pois $BA(x, y) = B(x, y, 0) = (x, y)$ para todo $(x, y) \in \mathbb{R}^2$.

Teorema 6.5. *Sejam E e F espaços vetoriais de dimensão finita. A transformação linear $A\colon E \to F$ possui inversa à esquerda se, e somente se, é injetiva.*

Demonstração: Seja $B\colon F \to E$ inversa à esquerda de A. Então $Au = Av \Rightarrow u = B(Au) = B(Av) = v$, logo A é injetiva. Reciprocamente, suponhamos que A seja injetiva. A fim de obter uma inversa à esquerda B para A, tomemos $\{v_1, \ldots, v_n\} \subset E$, uma base. Pelo Teorema 6.3, os vetores $Av_1, \ldots, Av_n \in F$ são L.I., logo podemos achar vetores $w_1, \ldots, w_k \in F$ tais que

$$\{Av_1, \ldots, Av_n, w_1, \ldots, w_k\} \subset F$$

seja uma base. (Teorema 3.4.) Pelo Teorema 4.1, a fim de definir a transformação linear $B\colon F \to E$, basta especificar seus valores nos

elementos desta base. Poremos $B(Av_1) = v_1, \ldots, B(Av_n) = v_n$, $Bw_1 = 0, \ldots, Bw_k = 0$. Dado qualquer $v \in E$, tem-se $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$, logo

$$\begin{aligned} BAv &= B(\alpha_1 Av_1 + \cdots + \alpha_n Av_n) \\ &= \alpha_1 BAv_1 + \cdots + \alpha_n BAv_n \\ &= \alpha_1 v_1 + \cdots + \alpha_n v_n = v, \end{aligned}$$

portanto B é uma inversa à esquerda de A. □

Uma transformação linear $A: E \to F$ chama-se *invertível* quando existe $B: F \to E$ linear tal que $BA = I_E$ e $AB = I_F$, ou seja, quando B é, ao mesmo tempo, inversa à esquerda e à direita de A.

Neste caso, diz-se que B é a *inversa* de A e escreve-se $B = A^{-1}$.

A fim de que a transformação linear A seja invertível, é necessário e suficiente que ela seja injetiva e sobrejetiva. Diz-se, então, que A é uma *bijeção linear* entre E e F ou, mais apropriadamente, que $A: E \to F$ é um *isomorfismo* e que os espaços vetoriais E e F são *isomorfos*.

Se $A: E \to F$ e $B: F \to G$ são isomorfismos, então $A^{-1}: F \to E$ e $BA: E \to G$ também são isomorfismos. Tem-se $(BA)^{-1} = A^{-1}B^{-1}$ e, para $\alpha \neq 0$, $(\alpha A)^{-1} = \frac{1}{\alpha} \cdot A^{-1}$.

Um isomorfismo $A: E \to F$ entre espaços vetoriais transforma toda base de E numa base de F. Reciprocamente, se uma transformação linear $A: E \to F$ leva alguma base de E numa base de F então A é um isomorfismo.

Do que foi dito acima resulta, em particular, que dois espaços vetoriais de dimensão finita isomorfos têm a mesma dimensão. A recíproca é verdadeira, como veremos agora.

Com efeito, seja E um espaço vetorial de dimensão finita n. Fixando uma base $\{v_1, \ldots, v_n\} \subset E$, podemos definir uma transformação linear $A: \mathbb{R}^n \to E$ pondo, para cada $v = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$, $Av = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Tem-se $Ae_1 = v_1, \ldots, Ae_n = v_n$. Assim, A transforma a base canônica $\{e_1, \ldots, e_n\} \subset \mathbb{R}^n$ na base $\{v_1, \ldots, v_n\} \subset E$, logo é um isomorfismo entre $\mathbb{R}^n$ e E.

Noutras palavras, todo espaço vetorial de dimensão n é isomorfo a $\mathbb{R}^n$.

Como o inverso $A^{-1}: E \to \mathbb{R}^n$ e o produto $BA^{-1}: E \to F$ de A por outro isomorfismo $B: \mathbb{R}^n \to F$ são isomorfismos, segue-se que dois espaços vetoriais E, F, ambos de dimensão n, são isomorfos.

Exemplo 6.8. O espaço $\mathcal{P}_n$, dos polinômios de grau $\leq n$, tem dimensão $n+1$, logo é isomorfo a $\mathbb{R}^{n+1}$. Por sua vez, o espaço $M(m\times p)$, das matrizes $m \times p$, é isomorfo a $\mathbb{R}^{mp}$, portanto $\mathcal{P}_n$ é isomorfo a $M(m \times p)$ se, e somente se, $n + 1 = mp$.

A noção de isomorfismo entre espaços vetoriais é fundamental .

Ela nos permite identificar, sob o ponto de vista da Álgebra Linear, espaços vetoriais que se apresentam sob formas à primeira vista diferentes. Por exemplo, a correspondência $(a_o, \ldots, a_n) \leftrightarrow a_o + a_1x + \cdots + a_nx^n$, é um isomorfismo natural entre $\mathbb{R}^{n+1}$ e $\mathcal{P}_n$, que desempenha papel relevante em Análise.

Noutro exemplo, se dispusermos os elementos de uma matriz $n \times n$ em fileiras paralelas à sua diagonal principal, veremos que há um total de

$$n + (n-1) + \cdots + 2 + 1 = \frac{n(n+1)}{2}$$

elementos nesta diagonal ou acima dela. Colocando esses elementos numa linha, em ordem determinada, obtemos um vetor de $\mathbb{R}^{n(n+1)/2}$. Se a matriz dada é simétrica, os elementos abaixo da diagonal não são necessários para determiná-la pois apenas repetem os demais. Este processo estabelece um isomorfismo entre o espaço vetorial S das matrizes simétricas $n \times n$ e o espaço euclidiano $\mathbb{R}^{n(n+1)/2}$, o que nos permite concluir que $\dim S = n(n+1)/2$. Um isomorfismo análogo (desprezando a diagonal principal que, neste caso, só contém zeros) mostra que as matrizes anti-simétricas $n \times n$ formam um espaço vetorial de dimensão $n(n-1)/2$.

Teorema 6.6. **(Teorema do Núcleo e da Imagem.)** *Sejam* E*,* F *espaços vetoriais de dimensão finita. Para toda transformação linear* $A\colon E \to F$ *tem-se* $\dim E = \dim \mathcal{N}(A) + \dim \mathcal{I}m(A)$.

Demonstração: O teorema resulta imediatamente da seguinte afirmação mais precisa, que provaremos a seguir: se $\{Au_1, \ldots, Au_p\}$ é uma base de $\mathcal{I}m(A)$ e $\{v_1, \ldots, v_q\}$ é uma base de $\mathcal{N}(A)$ então $\{u_1, \ldots, u_p, v_1, \ldots, v_q\}$ é uma base de E.

Com efeito, em primeiro lugar, se tivermos

$$\alpha_1 u_1 + \cdots + \alpha_p u_p + \beta_1 v_1 + \cdots + \beta_q v_q = 0 \qquad (*)$$

então, aplicando o operador A a ambos os membros desta igualdade e lembrando que $v_1, \ldots, v_q$ pertencem ao núcleo de A, obtemos

$$\alpha_1 Au_1 + \cdots + \alpha_p Au_p = 0.$$

Como os vetores $Au_1, \ldots, Au_p$ são L.I., resulta daí que $\alpha_1 = \cdots = \alpha_p = 0$. Portanto a igualdade (*) se reduz a

$$\beta_1 v_1 + \cdots + \beta_q v_q = 0.$$

Como $v_1, \ldots, v_q$ são L.I., concluímos que $\beta_1 = \cdots = \beta_q = 0$. Isto mostra que os vetores $u_1, \ldots, u_p$, $v_1, \ldots, v_q$ são L.I..

Em seguida, consideremos um vetor arbitrário $w \in E$. Como $Aw \in \mathcal{I}m(A)$, podemos escrever

$$Aw = \alpha_1\, Au_1 + \cdots + \alpha_p\, Au_p\,,$$

pois $\{Au_1, \ldots, Au_p\}$ é uma base da imagem de A. A igualdade acima pode ser reescrita como

$$A[w - (\alpha_1 u_1 + \cdots + \alpha_p u_p)] = 0.$$

Assim, o vetor $w - (\alpha_1 u_1 + \cdots + \alpha_p u_p)$ pertence ao núcleo de A, logo pode ser expresso como combinação linear dos elementos da base $\{v_1, \ldots, v_q\}$. Temos então

$$w - (\alpha_1 u_1 + \cdots + \alpha_p u_p) = \beta_1 v_1 + \cdots + \beta_q v_q\,,$$

ou seja, $w = \alpha_1 u_1 + \cdots + \alpha_p u_p + \beta_1 v_1 + \cdots + \beta_q v_q$. Isto mostra que os vetores $u_1, \ldots, u_p$, $v_1, \ldots, v_q$ geram E e portanto constituem uma base. □

Corolário. *Sejam* E, F *espaços vetoriais de mesma dimensão finita* n. *Uma transformação linear* $A\colon E \to F$ *é injetiva se, e somente se, é sobrejetiva e portanto é um isomorfismo.*

Com efeito, temos $n = \dim \mathcal{N}(A) + \dim \mathcal{I}m(A)$. Logo $\mathcal{N}(A) = \{0\}$ se, e somente se $\dim \mathcal{I}m(A) = n$, ou seja, $\mathcal{I}m(A) = F$. □

Exemplo 6.9. Um caso particular do corolário acima diz que, num espaço vetorial de dimensão finita, um operador linear é injetivo se, e somente se, é sobrejetivo. Isto seria falso num espaço de dimensão infinita, como se vê no seguinte exemplo: sejam $A, B\colon \mathbb{R}^\infty \to \mathbb{R}^\infty$ definidos por

$$A(x_1, x_2, x_3, \ldots) = (0, x_1, x_2, x_3, \ldots)$$

e

$$B(x_1, x_2, x_3, \ldots) = (x_2, x_3, x_4, \ldots).$$

A e B são operadores lineares. O primeiro é injetivo mas não é sobrejetivo e o segundo é sobrejetivo mas não é injetivo.

Exemplo 6.10. O Teorema do Núcleo e da Imagem dá outra explicação para o fato de um hiperplano $H \subset \mathbb{R}^n$ ter dimensão $n-1$. Por esse teorema, se $\dim E = n$ e $f\colon E \to \mathbb{R}$ é um funcional linear $\neq 0$ então o núcleo de f é um subespaço vetorial de dimensão $n-1$ em E, pois f não-nulo implica $\mathcal{I}m(f) = \mathbb{R}$ logo $\dim \mathcal{I}m(f) = 1$ e $\dim \mathcal{N}(f) = \dim E - \dim \mathcal{I}m(f) = n-1$. Ora, o hiperplano

$$H = \{(x_1, \ldots, x_n) \in \mathbb{R}^n; a_1x_1 + \cdots + a_nx_n = 0\}$$

é o núcleo do funcional linear não nulo $f\colon \mathbb{R}^n \to \mathbb{R}$, definido por

$$f(x_1, \ldots, x_n) = a_1x_1 + \cdots + a_nx_n.$$

Teorema 6.7. *Se uma transformação linear* $A\colon E \to F$ *tem uma inversa à esquerda* $B\colon F \to E$ *e uma inversa à direita* $C\colon F \to E$ *então* $B = C$ *e* A *é um isomorfismo, com* $A^{-1} = B = C$.

Demonstração: Tem-se $BA = I_E$ e $AC = I_F$. Portanto $B = BI_F = B(AC) = (BA)C = I_EC = C$. □

Corolário. *Seja* $\dim E = \dim F$. *Se as transformações lineares* $A\colon E \to F$, $B\colon F \to E$ *são tais que* $BA = I_E$ *então* $AB = I_F$ *e* $B = A^{-1}$.

Com efeito, $BA = I_E \Rightarrow A$ injetiva $\Rightarrow A$ sobrejetiva (Corolário do Teorema 6.6) $\Rightarrow AC = I_F$ para alguma $C \Rightarrow C = B$ (Teor. 6.7) $\Rightarrow AB = I_F$. □

Exercícios

6.1. Prove que o núcleo e a imagem de uma transformação linear $A\colon E \to F$ são respectivamente subespaços vetoriais de E e F.

6.2. Seja $A\colon E \to E$ um operador linear. Para quaisquer vetores $u \in \mathcal{N}(A)$ e $v \in \mathcal{I}m(A)$, prove que se tem $Au \in \mathcal{N}(A)$ e $Av \in \mathcal{I}m(A)$.

6.3. Encontre números a, b, c, d de modo que o operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, dado por $A(x,y) = (ax+by, cx+dy)$ tenha como núcleo a reta $y = 3x$.

6.4. Ache a, b, c, d tais que o operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ com $A(x,y) = (ax+by, cx+dy)$, tenha a reta $y = 2x$ como imagem.

6.5. Escreva a expressão de um operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ cujo núcleo seja a reta $y = x$ e cuja imagem seja a reta $y = 2x$.

6.6. Defina um operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ que tenha como núcleo e como imagem o eixo x.

6.7. Resolva um exercício análogo ao anterior, com a reta $y = 5x$ em lugar do eixo x.

6.8. Considere a transformação linear $A\colon \mathbb{R}^4 \to \mathbb{R}^3$, dada por

$$A(x,y,z,t) = (x+y+z+2t, x-y+2z, 4x+2y+5z+6t),$$

encontre um vetor $b \in \mathbb{R}^3$ que não pertença à imagem de A e com isso exiba um sistema linear de três equações com quatro incógnitas sem solução.

6.9. Seja $E = C^0(\mathbb{R})$ o espaço das funções contínuas $f\colon \mathbb{R} \to \mathbb{R}$. Defina o operador linear $A\colon E \to E$ pondo, para cada $f \in E$, $Af = \varphi$, onde $\varphi(x) = \int_0^x f(t)\,dt$, $x \in \mathbb{R}$. Determine o núcleo e a imagem do operador A.

6.10. Seja $E = \mathbb{R}^\infty$ o espaço vetorial cujos elementos são as seqüências $x = (x_1, x_2, \ldots)$ de números reais. Defina os operadores lineares $A, B\colon E \to E$ pondo $Ax = (x_1, 0, x_2, 0, x_3, \ldots)$ e $Bx = y$, onde $y = (y_1, y_2, \ldots)$, onde $y_k = x_{k+1} - 2x_k$. Determine o núcleo e a imagem de A e B.

6.11. Assinale verdadeiro (V) ou falso (F):

( ) Uma transformação linear $A\colon E \to F$ é sobrejetiva se, e somente se, $\dim \mathcal{N}(A) = \dim E - \dim F$.

( ) Dada a transformação linear $A\colon E \to F$, para todo b fixado em F, o conjunto $G = \{x \in E; Ax = b\}$ é um subespaço vetorial de E.

( ) Para todo operador linear $A\colon E \to E$, tem-se $E = \mathcal{N}(A) \oplus \mathcal{I}m\,A$.

( ) Todo operador linear injetivo no espaço $C^0(\mathbb{R})$ das funções contínuas $f\colon \mathbb{R} \to \mathbb{R}$ é também sobrejetivo.

( ) O núcleo de toda transformação linear $A\colon \mathbb{R}^5 \to \mathbb{R}^3$ tem dimensão ≥ 3.

6.12. Seja $\mathbf{a} = [a_{ij}]$ uma matriz $m \times n$. Se suas colunas geram um subespaço vetorial F, de dimensão r, em $\mathbb{R}^m$ prove que, para todo $b \in F$, as soluções $x = (x_1, \ldots, x_n)$ do sistema linear $\sum_{j=1}^{n} a_{ij}x_j = b_i$ $(i = 1, \ldots, m)$ formam uma variedade afim de dimensão $n - r$ em $\mathbb{R}^n$.

6.13. Prove que cada uma das transformações lineares abaixo é injetiva e obtenha uma inversa à esquerda linear para cada uma delas.

(a) $A\colon \mathbb{R} \to \mathbb{R}^n$; $A(x) = (x, 2x, \ldots, nx)$.

(b) $B\colon \mathbb{R}^2 \to \mathbb{R}^3$; $B(x, y) = (x + 2y, x + y, x - y)$.

(c) $D\colon \mathbb{R}^3 \to \mathbb{R}^4$; $D(x, y, z) = (2x, 3y, 5z, x + y + z)$.

(d) $C\colon \mathcal{P}_n \to \mathcal{P}_{n+2}$; $C \cdot p(x) = (x^2 + 1)p(x)$.

6.14. Seja E um espaço vetorial de dimensão finita. Dado um operador linear $A\colon E \to E$, defina o novo operador $T_A\colon \mathcal{L}(E) \to \mathcal{L}(E)$ pondo $T_A(X) = AX$, para todo $X \in \mathcal{L}(E)$. Prove que T_A é invertível se, e somente se, A é invertível. Mesmo problema para $S_A(X) = XA$.

6.15. Sejam F_1, F_2 subespaços de E tais que $\dim F_1 + \dim F_2 = \dim E$. Mostre que existe um operador linear $A\colon E \to E$ tal que $F_1 = \mathcal{N}(A)$ e $F_2 = \mathcal{I}m(A)$.

6.16. Prove que uma transformação linear $A\colon E \to F$ é sobrejetiva se, e somente se, transforma um conjunto de geradores de E num conjunto de geradores de F.

6.17. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ um operador linear $\neq 0$. Se $A^n = 0$ para algum $n > 2$, prove que $A^2 = 0$. [Sugestão: seja $F = \mathcal{I}m\,A$. Então a restrição de A à reta F é zero ou é invertível.]

6.18. Seja $A\colon \mathcal{P}_n \to \mathcal{P}_n$ o operador linear definido por $A \cdot p(x) = x \cdot p'''(x)$. Descreva o núcleo e a imagem de A. Obtenha bases para $\mathcal{N}(A)$ e para $\mathcal{I}m(A)$.

6.19. Assinale verdadeiro (V) ou falso (F):

( ) Se a transformação linear $A \colon \mathbb{R}^m \to \mathbb{R}^n$ é injetiva então $\dim \mathcal{I}m(A) = m$.

( ) Se $A \colon \mathbb{R}^m \to \mathbb{R}^n$ é sobrejetiva então $\dim \mathcal{N}(A) = m - n$.

( ) Se $A, B \in \mathcal{L}(\mathbb{R}^n)$ são tais que $\dim \mathcal{I}m(A) = \dim \mathcal{I}m(B)$ então $\dim \mathcal{I}m(AB) = \dim \mathcal{I}m(BA) = \dim \mathcal{I}m(A)$.

( ) Se $\mathcal{N}(A)$ é gerado pelos vetores v_1, v_2, v_3 então a imagem do operador $A \colon \mathbb{R}^5 \to \mathbb{R}^5$ tem dimensão 2.

6.20. Determine uma base para a imagem de cada uma das transformações lineares abaixo e indique quais são sobrejetivas.

(a) $A \colon \mathbb{R}^2 \to \mathbb{R}^2$, $A(x, y) = (x - y, x - y)$.

(b) $B \colon \mathbb{R}^4 \to \mathbb{R}^4$, $B(x, y, z, t) = (x + y, z + t, x + z, y + t)$.

(c) $C \colon \mathbb{R}^3 \to \mathbb{R}^3$, $C(x, y, z) = \left(x + \frac{y}{2}, y + \frac{z}{2}, z + \frac{x}{2}\right)$.

(d) $D \colon M(2 \times 2) \to M(2 \times 2)$, $D \cdot X = AX$, onde $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

(e) $E \colon \mathcal{P}_n \to \mathcal{P}_{n+1}$, $E \cdot p(x) = x \cdot p(x)$.

6.21. Prove que as transformações lineares a seguir são sobrejetivas e obtenha uma inversa à direita linear para cada uma delas.

(a) $A \colon \mathbb{R}^3 \to \mathbb{R}^2$, $A(x, y, z) = (2x + y, z)$.

(b) $B \colon \mathcal{P}_n \to \mathbb{R}$, $B \cdot p(x) = p(1)$.

(c) $C \colon \mathbb{R}^2 \to \mathbb{R}^2$, $C(x, y) = (x + y, x - y)$.

(d) $P \colon \mathbb{R}^n \to \mathbb{R}^{n-1}$, $P(x_1, \ldots, x_n) = (x_1, \ldots, x_{n-1})$.

Mostre que em três dos quatro exemplos acima, as transformações dadas admitem infinitas inversas à direita lineares e infinitas não-lineares.

6.22. Para cada uma das nove transformações lineares dos exercícios anteriores, determine o núcleo e obtenha uma base do mesmo, caso não se reduza a $\{0\}$.

6.23. Seja $T\colon \mathcal{P}_n \to \mathcal{P}_n$ o operador linear definido por $T.p(x) = 5p(x) - 4p'(x) + p''(x)$. Mostre que seu núcleo é $\{0\}$ e conclua que, para todo polinômio $b(x)$ existe um polinômio $p(x)$ tal que $5p(x) - 4p'(x) + p''(x) = b(x)$.

6.24. Seja $X \subset F$ um subconjunto com a seguinte propriedade: toda transformação linear $A\colon E \to F$ cuja imagem contém X é sobrejetiva. Prove que X é um conjunto de geradores de F.
(Sugestão: é muito mais fácil do que parece.)

6.25. Seja $A\colon E \to F$ uma transformação linear sobrejetiva entre espaços vetoriais de dimensão finita. Prove que existe um subespaço vetorial $E' \subset E$ tal que a restrição de A a E' é um isomorfismo sobre F.

6.26. Seja $A\colon E \to F$ um isomorfismo. Se $T\colon F \to E$ é tal que $AT = I_F$ (ou $TA = I_E$), prove que T é linear.

6.27. Sejam E, F espaços vetoriais de dimensão finita. Se $\dim E < \dim F$, prove que existem transformações lineares $A\colon E \to F$ e $B\colon F \to E$ tais que A é injetiva e B é sobrejetiva.

6.28. Dadas as transformações lineares $A\colon E \to F$, $B\colon F \to G$, assinale V(erdadeiro) ou F(also) nas seguintes implicações:

( ) BA sobrejetiva $\Rightarrow$ B sobrejetiva.
( ) BA sobrejetiva $\Rightarrow$ A sobrejetiva.
( ) BA injetiva $\Rightarrow$ B injetiva.
( ) BA injetiva $\Rightarrow$ A injetiva.

Prove ainda que se $E = F = G$ então as quatro implicações são verdadeiras.

6.29. Prove que toda transformação linear A pode escrever-se como o produto $A = TS$, onde S é uma transformação linear sobrejetiva e T é uma transformação linear injetiva. Vale também uma decomposição do tipo $A = S'T'$, com S' sobrejetiva e T' injetiva?

6.30. Dado o conjunto finito X com n elementos, prove que o espaço vetorial $\mathcal{F}(X;\mathbb{R})$ é isomorfo a $\mathbb{R}^n$. Mais geralmente, dado qualquer

espaço vetorial E, estabeleça um isomorfismo entre $\mathcal{F}(X;E)$ e $E^n = E \times \cdots \times E$ (n fatores).

6.31. Dados os números reais $a = a_o < a_1 < \cdots < a_n = b$, considere o conjunto $E_1 \subset \mathcal{F}([a,b];\mathbb{R})$ formado pelas funções $f\colon [a,b] \to \mathbb{R}$ que, em cada intervalo fechado $[a_{i-1}, a_i]$, $i = 1, \ldots, n$, são representadas por polinômios de grau ≤ 1, isto é, $f(x) = \alpha_i x + \beta_i$ para $a_{i-1} \leq x \leq a_i$. Prove que E_1 é um subespaço vetorial e que a correspondência $f \mapsto (f(a_o), \ldots, f(a_n))$ é um isomorfismo entre E_1 e $\mathbb{R}^{n+1}$. Conclua que $\dim E_1 = n+1$. Descreva a base de E_1 que corresponde à base canônica de $\mathbb{R}^{n+1}$.

6.32. Com a notação de exercício anterior, seja E_2 o conjunto das funções deriváveis $f\colon [a,b] \to \mathbb{R}$ cujas restrições aos intervalos $[a_{i-1}, a_i]$, $i = 1, \ldots, n$, são polinômios de grau ≤ 2. Considerando a derivação $D\colon E_2 \to E_1$, prove que $E_2 \subset \mathcal{F}([a,b];\mathbb{R})$ é um subespaço vetorial de dimensão $n+2$.

6.33. Sejam E um espaço vetorial de dimensão finita, $E^* = \mathcal{L}(E;\mathbb{R})$ seu dual e $E^{**} = \mathcal{L}(E^*;\mathbb{R})$ seu *bi-dual*. Considere a correspondência $\xi\colon E \to E^{**}$, que associa a cada vetor $v \in E$ o elemento $\xi(v) = v^{**} \in E^{**}$ tal que $v^{**}(f) = f(v)$ para todo $v \in E$. Prove que ξ é um isomorfismo. [Este exercício significa que $f(v)$ pode ser considerado como um valor da função f de variável v ou da função v (mais exatamente, v^{**}) de variável f.]

6.34. Seja $f\colon E \to \mathbb{R}$ um funcional linear não-nulo no espaço vetorial E, de dimensão n. Prove que existe uma base $\{u_1, \ldots, u_n\} \subset E$ tal que $f(u_1) = \cdots = f(u_{n-1}) = 0$ e $f(u_n) = 1$. Use este fato para provar que se $g\colon E \to \mathbb{R}$ é outro funcional linear não-nulo então existe um isomorfismo $A\colon E \to E$ tal que $g = f \circ A$.

6.35. Dado o funcional linear não-nulo $f\colon E \to \mathbb{R}$, prove que existe um vetor $u \in E$ tal que $f(u) = 1$. Seja $F \subset E$ o subespaço (reta) gerado por u. Prove que $E = F \oplus \mathcal{N}(f)$.

6.36. Sejam $f, g\colon E \to \mathbb{R}$ funcionais lineares não-nulos no espaço vetorial E, de dimensão finita. Prove que um deles é múltiplo do outro se, e somente se, eles têm o mesmo núcleo.

6.37. Supondo $Y \subset X$, descreva o núcleo e a imagem da transformação linear $R\colon \mathcal{F}(X;\mathbb{R}) \to \mathcal{F}(Y;\mathbb{R})$, que associa a cada função $f\colon X \to \mathbb{R}$ sua restrição $Rf = f|Y\colon Y \to \mathbb{R}$.

6.38. Sejam E, F os espaços vetoriais cujos elementos são respectivamente as funções pares e as funções ímpares $\mathbb{R} \to \mathbb{R}$. Descreva o núcleo e a imagem das transformações lineares $R\colon E \to \mathcal{F}([0,+\infty);\mathbb{R})$ e $R'\colon F \to \mathcal{F}([0,+\infty);\mathbb{R})$ que associam a cada função $f\colon \mathbb{R} \to \mathbb{R}$ sua restrição ao intervalo $[0,+\infty)$.

6.39. Estabeleça um isomorfismo entre o espaço vetorial das matrizes simétricas $n \times n$ e o espaço das matrizes triangulares inferiores ($a_{ij} = 0$ se $i < j$). Idem entre as matrizes anti-simétricas e as triangulares inferiores com diagonal nula.

6.40. Sejam F_1 e F_2 subespaços vetoriais de dimensão 3 em $\mathbb{R}^5$. Quais são as dimensões possíveis do subespaço $F_1 \cap F_2$? Mesma pergunta com $\dim F_1 = 4$ e $\dim F_2 = 3$.

6.41. Seja $A\colon E \to E$ um operador linear. Prove que $A^2 = 0$ se, e somente se, $\mathcal{I}m(A) \subset \mathcal{N}(A)$.

6.42. Dadas as transformações lineares $A, B\colon E \to F$, entre espaços vetoriais de dimensão finita, prove:

(a) Se $\mathcal{N}(A) = \mathcal{N}(B)$ então existe um isomorfismo $Q\colon F \to F$ tal que $B = QA$.

(b) Se $\mathcal{I}m(A) = \mathcal{I}m(B)$ então existe um isomorfismo $P\colon E \to E$ tal que $B = AP$.

(c) Se $\dim \mathcal{N}(A) = \dim \mathcal{N}(B)$ (ou, equivalentemente, $\dim \mathcal{I}m(A) = \dim \mathcal{I}m(B)$) então existem isomorfismos $P\colon E \to E$ e $Q\colon F \to F$ tais que $B = QAP$.

(d) Existem operadores lineares $A, B\colon E \to E$ tais que $\mathcal{N}(A)=\mathcal{N}(B)$, $\mathcal{I}m(A) = \mathcal{I}m(B)$, sem que exista um isomorfismo $P\colon E \to E$, com $B = P^{-1}AP$.

6.43. Se os vetores $v_1, \ldots, v_m \in E$ geram um subespaço vetorial de dimensão r, prove que o conjunto dos vetores $(\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$ tais que $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$ é um subespaço vetorial de $\mathbb{R}^m$ com dimensão $m - r$. [Sugestão: considere a transformação linear $(\alpha_1, \ldots, \alpha_m) \mapsto \alpha_1 v_1 + \cdots + \alpha_m v_m$, de $\mathbb{R}^m$ em E.]

6.44. Seja $A\colon E \to E$ um operador linear tal que $A^k = 0$ para algum número natural k. Prove que, para todo $\alpha \neq 0$, o operador $A - \alpha I$ é invertível.

6.45. Se, para algum $\alpha_o \neq 0$, tem-se $\alpha_o I + \alpha_1 A + \cdots + \alpha_m A^m = 0$, prove que o operador linear A é invertível.

6.46. Sem fazer hipóteses sobre as dimensões de E e F, sejam $A\colon E \to F$ e $B\colon F \to E$ transformações lineares. Se AB é invertível, prove que A é sobrejetiva e B é injetiva. Se AB e BA são invertíveis, prove que A é invertível.

6.47. Calcule $(B^{-1}AB)^m$.

7

Soma Direta e Projeção

Esta seção trata da decomposição de um espaço vetorial como soma de subespaços independentes, mostra que essa decomposição equivale a definir um operador idempotente no espaço e estabelece a conexão entre projeções e involuções, ou simetrias.

Na Seção 2, vimos que se F_1 e F_2 são subespaços do espaço vetorial E, o subespaço vetorial de E gerado pela reunião $F_1 \cup F_2$ é o conjunto $F_1 + F_2$ de todas as somas $u + v$, onde $u \in F_1$ e $v \in F_2$. No caso particular em que $F_1 \cap F_2 = \{0\}$, escreve-se $F_1 \oplus F_2$ em vez de $F_1 + F_2$, diz-se que $F_1 \oplus F_2$ é a *soma direta* de F_1 com F_2 e prova-se (Teorema 2.1) que a condição $F_1 \cap F_2 = \{0\}$ equivale a dizer que $u + v = u' + v'$, com $u, u' \in F_1$ e $v, v' \in F_2$, implica $u = u'$ e $v = v'$.

Existe uma noção análoga à de soma direta, que é o *produto cartesiano* $E_1 \times E_2$ de dois espaços vetoriais E_1 e E_2. Aqui E_1 e E_2 não precisam ser subespaços vetoriais do mesmo espaço E. Os elementos do conjunto $E_1 \times E_2$ são os pares ordenados (u, v), onde $u \in E_1$ e $v \in E_2$. As operações que tornam $E_1 \times E_2$ um espaço vetorial são definidas por

$$(u, v) + (u', v') = (u + u', v + v'), \quad \alpha(u, v) = (\alpha u, \alpha v),$$

para quaisquer $u, u' \in E_1$, $v, v' \in E_2$ e $\alpha \in \mathbb{R}$. O vetor nulo de $E_1 \times E_2$ é o par (0,0) e o inverso aditivo de (u, v) é $(-u, -v)$. Se

$\{u_1, \ldots, u_m\} \subset E_1$ e $\{v_1, \ldots, v_n\} \subset E_2$ são bases, é imediato constatar que $\{(u_1, 0), \ldots, (u_m, 0), (0, v_1), \ldots, (0, v_n)\} \subset E_1 \times E_2$ é uma base, de modo que $\dim(E_1 \times E_2) = \dim E_1 + \dim E_2$.

Se F_1 e F_2 são subespaços vetoriais de E, com $F_1 \cap F_2 = \{0\}$, então a transformação linear

$$A \colon F_1 \times F_2 \to F_1 \oplus F_2,$$

definida por $A(u, v) = u + v$, $u \in F_1$, $v \in F_2$, é um isomorfismo, como se verifica facilmente. Se $\{u_1, \ldots, u_m\} \subset F_1$ e $\{v_1, \ldots, v_n\} \subset F_2$ são bases então a base $\{(u_1, 0), \ldots, (u_m, 0), (0, v_1), \ldots (0, v_n)\}$ de $F_1 \times F_2$ é transformada por A no conjunto $\{u_1, \ldots, u_m, v_1, \ldots, v_n\}$ o qual é, por conseguinte, uma base de $F_1 \oplus F_2$. Segue-se que $\dim(F_1 \oplus F_2) = \dim F_1 + \dim F_2 = m + n$.

No caso mais geral, em que a interseção $F_1 \cap F_2$ dos dois subespaços $F_1, F_2 \subset E$ não se reduz necessariamente ao vetor nulo, a soma $F_1 + F_2$ pode não ser uma soma direta mas, mesmo assim, está bem definida a transformação linear

$$A \colon F_1 \times F_2 \longrightarrow F_1 + F_2,$$

onde $A(u, v) = u + v$ para $u \in F_1$ e $v \in F_2$. Obviamente, A é sobrejetiva. Seu núcleo é formado pelos pares (u, v) tais que $u + v = 0$, isto é, $v = -u$, logo u e v pertencem ambos a F_1 e a F_2. Noutras palavras, $\mathcal{N}(A) = \{(u, -u); u \in F_1 \cap F_2\}$. A correspondência $u \mapsto (u, -u)$ é um isomorfismo evidente entre $F_1 \cap F_2$ e $\mathcal{N}(A)$. Pelo Teorema do Núcleo e da Imagem, temos

$$\begin{aligned} \dim F_1 + \dim F_2 &= \dim(F_1 \times F_2) \\ &= \dim \mathcal{N}(A) + \dim(F_1 + F_2) \\ &= \dim(F_1 \cap F_2) + \dim(F_1 + F_2). \end{aligned}$$

Isto nos permite enunciar o

Teorema 7.1. *Sejam F_1 e F_2 subespaços de dimensão finita de um espaço vetorial E. Tem-se* $\dim F_1 + \dim F_2 = \dim(F_1 \cap F_2) + \dim(F_1 + F_2)$.

A noção de soma direta está intimamente ligada à noção de projeção. Se $E = F_1 \oplus F_2$ é a decomposição do espaço vetorial E como soma

direta dos subespaços F_1 e F_2, define-se o operador linear $P\colon E \to E$, *projeção de* E *sobre* F_1*, paralelamente a* F_2, do seguinte modo: todo vetor $w \in E$ se escreve, de modo único, como soma $w = u + v$ de um vetor $u \in F_1$ com um vetor $v \in F_2$. Põe-se então $Pw = u$. (Veja Fig. 7.1.)

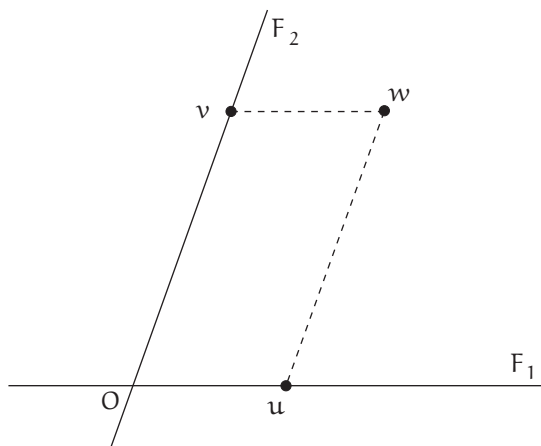



Figura 7.1.

O operador linear $P\colon E \to E$ assim definido tem imagem F_1 e núcleo F_2. Além disso, como se vê facilmente, P é *idempotente*, isto é, $P^2 = P$. O teorema seguinte mostra que, reciprocamente, todo operador linear idempotente é uma projeção.

Preliminarmente, observemos que se $P^2 = P$ então, para todo $w \in \mathcal{I}m(P)$, tem-se $Pw = w$ pois $w \in \mathcal{I}m(P) \Rightarrow w = Pv \Rightarrow Pw = PPv = Pv = w$.

Teorema 7.2. *Seja* $P\colon E \to E$ *um operador linear. Se* $P^2 = P$ *então* E *é a soma direta do núcleo com a imagem de* P*. Além disso,* P *é a projeção sobre* $\mathcal{I}m(P)$ *paralelamente a* $\mathcal{N}(P)$.

Demonstração: Todo $v \in E$ escreve-se como soma $v = (v - Pv) + Pv$, onde Pv, evidentemente, pertence a $\mathcal{I}m(P)$ e, como $P(v - Pv) = Pv - PPv = Pv - Pv = 0$, vemos que $v - Pv \in \mathcal{N}(P)$. Portanto $E = \mathcal{N}(P) + \mathcal{I}m(P)$. Se $w \in \mathcal{N}(P) \cap \mathcal{I}m(P)$, por um lado tem-se $Pw = 0$ e, por outro, $Pw = w$; logo $w = 0$. Assim $\mathcal{N}(P) \cap \mathcal{I}m(P) = \{0\}$ e tem-se a soma direta $E = \mathcal{N}(P) \oplus \mathcal{I}m(P)$. A última afirmação do enunciado é óbvia. □

Exemplo 7.1. Para todo operador linear $A\colon E \to E$ num espaço vetorial de dimensão finita vale a relação $\dim E = \dim \mathcal{N}(A) + \dim \mathcal{I}m(A)$. Isto porém não implica que se tenha sempre $E = \mathcal{N}(A) \oplus \mathcal{I}m(A)$. Por exemplo, se $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ é definido por $A(x,y) = (x-y, x-y)$ então, tomando $w = (1,1)$, temos $w = Av$, com $v = (2,1)$ e $Aw = 0$, logo $\mathcal{N}(A) \cap \mathcal{I}m(A)$ contém o vetor não-nulo w.

Outro exemplo de operador linear que está ligado à decomposição de um espaço vetorial como soma direta de dois subespaços é fornecido pelas involuções.

Uma *involução* é um operador linear $S\colon E \to E$ tal que $S^2 = I$, ou seja, $S(Sv) = v$ para todo $v \in E$.

Noutras palavras, uma involução é um operador invertível, igual ao seu próprio inverso. Um exemplo de involução é a reflexão (ortogonal) no plano em torno de uma reta que passa pela origem.

Veremos agora que toda involução é a reflexão em torno de um subespaço, paralelamente a outro.

Teorema 7.3. *Seja* $S\colon E \to E$ *uma involução. Os conjuntos* $F_1 = \{u \in E; Su = u\}$ *e* $F_2 = \{v \in E; Sv = -v\}$ *são subespaços vetoriais e* $E = F_1 \oplus F_2$. *Para todo* $w = u+v$, *com* $u \in F_1$ *e* $v \in F_2$ *tem-se* $Sw = u-v$. *Além disso,* $P = \frac{1}{2}(S + I)$ *é a projeção sobre* F_1, *paralelamente a* F_2. (Veja Fig. 7.2.)

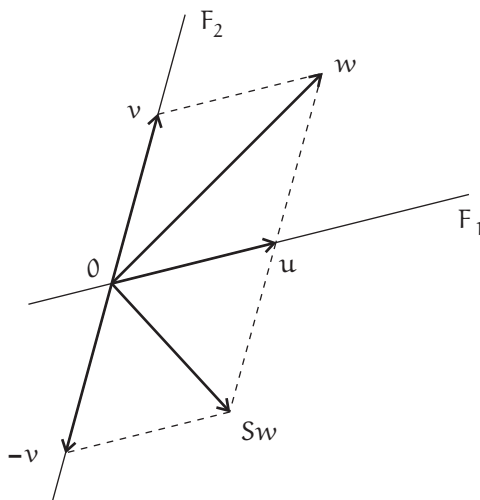



Figura 7.2.

Demonstração: Para todo $w \in E$, podemos escrever $w = u + v$, onde $u = (w + Sw)/2$ e $v = (w - Sw)/2$. Como $S^2 = I$, é claro que $Su = u$ e $Sv = -v$, ou seja, $u \in F_1$ e $v \in F_2$. É claro também que $F_1 \cap F_2 = \{0\}$ e que $w = u + v \Rightarrow Sw = u - v$ se $u \in F_1$ e $v \in F_2$. Finalmente,

$$\begin{aligned} P = \frac{1}{2}(S + I) \Rightarrow P^2 &= \frac{1}{4}(S^2 + 2S + I) \\ &= \frac{1}{4}(2S + 2I) \\ &= \frac{1}{2}(S + I) = P. \end{aligned}$$

Vê-se facilmente que o núcleo de P é F_2 e a imagem de P é F_1. □

Na situação descrita pelo Teorema 7.3, diz-se que a involução S é a reflexão em torno do subespaço F_1, paralelamente a F_2. O caso mais comum de reflexão é aquele em que se tem $\dim E = n$, $\dim F_1 = n-1$ e $\dim F_2 = 1$, de modo que S é a reflexão em torno do hiperplano F_1 paralelamente à reta F_2.

Exercícios

7.1. No plano $\mathbb{R}^2$, considere as retas F_1 e F_2, definidas respectivamente pelas equações $y = ax$ e $y = bx$, com $a \neq b$. Em seguida:

(1) Exprima cada vetor $v = (x, y) \in \mathbb{R}^2$ como soma de um vetor em F_1 e um vetor em F_2.

(2) Obtenha a matriz (em relação à base canônica) da projeção $P\colon \mathbb{R}^2 \to \mathbb{R}^2$, que tem F_1 como núcleo e F_2 como imagem.

(3) Ache a matriz da reflexão $S\colon \mathbb{R}^2 \to \mathbb{R}^2$, em torno da reta F_2, paralelamente a F_1.

7.2. Se $P, Q\colon E \to E$ são projeções e $PQ = QP$, prove que PQ é uma projeção cujo núcleo é $\mathcal{N}(P) + \mathcal{N}(Q)$ e cuja imagem é $\mathcal{I}m(P) \cap \mathcal{I}m(Q)$.

7.3. Exprima um vetor arbitrário $v = (x, y, z) \in \mathbb{R}^3$ como soma de um vetor do plano F_1, cuja equação é $x + y - z = 0$, com um vetor da reta

F_2, gerada pelo vetor $(1,2,1)$. Conclua que $\mathbb{R}^3 = F_1 \oplus F_2$. Determine a matriz (relativa à base canônica) da projeção $P\colon \mathbb{R}^3 \to \mathbb{R}^3$, que tem imagem F_1 e núcleo F_2.

7.4. É dado um operador linear $P\colon E \to E$. Assinale verdadeiro (V) ou falso (F):

( ) Se $E = \mathcal{N}(P) \oplus \mathcal{I}m(P)$ então P é uma projeção.

( ) Se $E = \mathcal{N}(P) + \mathcal{I}m(P)$ então P é uma projeção.

( ) Se P é uma projeção então $I - P$ também é.

( ) Se P é uma projeção então $\mathcal{I}m(P) = \mathcal{N}(I-P)$ e $\mathcal{N}(P) = \mathcal{I}m(I-P)$.

7.5. Se $\mathcal{N}(P) = \mathcal{I}m(I-P)$, prove que o operador linear $P\colon E \to E$ é uma projeção.

7.6. Mostre que

$$\begin{bmatrix} 1 & 0 & a & b \\ 0 & 1 & c & d \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

é a matriz (na base canônica) de uma projeção $P\colon \mathbb{R}^4 \to \mathbb{R}^4$. Escreva as equações que definem o núcleo e a imagem dessa projeção.

7.7. Prove que o operador $P\colon \mathbb{R}^2 \to \mathbb{R}^2$, dado por $P(x,y) = (-2x - 4y, \frac{3}{2}x + 3y)$ é a projeção sobre uma reta. Determine o núcleo e a imagem de P.

7.8. Considere o operador linear $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, dado por

$$A(x,y,z) = (40x + 18y - 6z, 18x + 13y + 12z, -6x + 12y + 45z).$$

Mostre que $P = \frac{1}{49} \cdot A$ é uma projeção, que $\mathcal{I}m(P)$ é um plano e determine a equação desse plano.

7.9. Sejam F_1, F_2 subespaços vetoriais de E, com $\dim F_1 + \dim F_2 = \dim E$ (dimensões finitas). Prove que $E = F_1 \oplus F_2$ se, e somente se, $F_1 \cap F_2 = \{0\}$.

7.10. Seja $A: E \to E$ um operador linear num espaço vetorial de dimensão finita. Prove que $E = \mathcal{N}(A) \oplus \mathcal{I}m(A)$ se, e somente se, $\mathcal{N}(A) = \mathcal{N}(A^2)$.

7.11. Suponha que o espaço vetorial de dimensão finita E admita a decomposição $E = F_1 \oplus \cdots \oplus F_k$, como soma direta de subespaços vetoriais. (Vide Exercício 2.33.) Para cada $i = 1, \ldots, k$, escreva $G_i = F_1 \oplus \cdots \oplus F_{i-1} \oplus F_{i+1} \oplus \cdots \oplus F_k$ e chame de $P_i: E \to E$ a projeção sobre F_i, paralelamente a G_i. Prove que $P_1 + \cdots + P_k = I$ e $P_iP_j = 0$ se $i \neq j$.

7.12. Sejam $P_1, \ldots, P_k: E \to E$ operadores lineares tais que $P_1 + \cdots + P_k = I$ e $P_iP_j = 0$ se $i \neq j$. Prove que esses operadores são projeções.

7.13. Sejam $P, Q: E \to E$ projeções. Prove que as seguintes afirmações são equivalentes:

(a) $P + Q$ é uma projeção;

(b) $PQ + QP = 0$;

(c) $PQ = QP = 0$.

[Para provar que (b) $\Rightarrow$ (c), multiplique à esquerda, e depois à direita, por P.]

7.14. Prove que o produto de duas involuções é uma involução se, e somente se, elas comutam.

7.15. Mostre que os seguintes operadores são involuções e determine, em cada caso, a projeção correspondente na forma do Teorema 7.3.

(a) $S: \mathcal{F}(\mathbb{R}^2; \mathbb{R}) \to \mathcal{F}(\mathbb{R}^2; \mathbb{R})$, $Sf = f^*$, $f^*(x, y) = f(y, x)$.

(b) $U: \mathcal{F}(\mathbb{R}^+; \mathbb{R}) \to \mathcal{F}(\mathbb{R}^+; \mathbb{R})$, $Uf = \hat{f}$, $\hat{f}(x) = f(1/x)$.

(c) $V: \mathbb{R}^n \to \mathbb{R}^n$, $V(x_1, \ldots, x_n) = (-x_1, \ldots, -x_k, x_{k+1}, \ldots, x_n)$.

7.16. Se o espaço vetorial E tem dimensão finita, prove que para todo subespaço $F \subset E$ existe (pelo menos) um subespaço $G \subset E$ tal que $E = F \oplus G$.

7.17. Seja $E = F_1 \oplus F_2$. O *gráfico* de uma transformação linear $A\colon F_1 \to F_2$ é o subconjunto $G \subset E$ formado pelas somas $v + Av$, onde $v \in F_1$. Prove que G é um subespaço vetorial de E e que a projeção $P\colon E \to F_1$, restrita a G, define um isomorfismo entre G e F_1. Reciprocamente, se $G \subset E$ é um subespaço vetorial tal que a restrição de P a G é um isomorfismo de G sobre F_1, prove que G é o gráfico de uma transformação linear $A\colon F_1 \to F_2$.

7.18. Diz-se que $X \cup Y = Z$ é uma *partição* de Z quando $X \cap Y = \emptyset$. Se $J \cup K = \{1, \dots, n\}$ é uma partição, prove que $\mathbb{R}^n = \mathbb{R}^J \oplus \mathbb{R}^K$, onde $\mathbb{R}^J$ e $\mathbb{R}^K$ são os subespaços vetoriais de $\mathbb{R}^n$ gerados pelos vetores e_j, $j \in J$ e pelos vetores e_k, $k \in K$ respectivamente. Seja $F \subset \mathbb{R}^n$ um subespaço vetorial. Prove que existe uma partição $J \cup K = \{1, \dots, n\}$ tal que F é o gráfico de uma transformação linear $A\colon \mathbb{R}^J \to \mathbb{R}^K$, onde $\dim F =$ número de elementos de J.

7.19. Seja $P\colon E \to E$ uma projeção. Prove que os vetores v e $(1-t)v + tPv$, para todo $v \in E$ e todo $t \in \mathbb{R}$, têm a mesma imagem por P.

7.20. Sejam $P, Q\colon E \to E$ projeções. Se $P + Q$ for ainda uma projeção, prove que $\operatorname{Im}(P+Q) = \operatorname{Im}(P) \oplus \operatorname{Im}(Q)$. Considerando as projeções $P, Q\colon \mathbb{R}^2 \to \mathbb{R}^2$, com $P(x,y) = (x,0)$ e $Q(x,y) = \frac{1}{2}(x+y, x+y)$, mostre que a recíproca é falsa.

7.21. Prove que todo espaço vetorial de dimensão finita é soma direta de subespaços de dimensão 1.

8

A Matriz de uma Transformação Linear

A matriz de uma transformação linear é um objeto concreto, associado a essa transformação na presença de bases em seu domínio e seu contra-domínio. A matriz permite obter uma variedade ilimitada de exemplos de transformações lineares, bem como calcular especificamente a imagem de um dado vetor por uma transformação. Nesta seção será estudada a relação entre uma transformação linear e sua matriz. Em particular, o produto de transformações conduzirá a uma profícua noção de produto de matrizes. Veremos como se relacionam as matrizes da mesma transformação tomadas em bases diferentes e daremos uma demonstração direta da igualdade entre o posto-linha e o posto-coluna de uma matriz.

Vimos na Seção 4 que uma transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ fica inteiramente determinada pela matriz $\mathbf{a} = [a_{ij}] \in M(m \times n)$, cujo ij-ésimo termo a_{ij} é a i-ésima coordenada do vetor $A \cdot e_j \in \mathbb{R}^m$. Com efeito, conhecendo essa matriz tem-se, para cada $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$, o valor $A \cdot v = (y_1, \ldots, y_m)$ dado por

$$y_i = a_{i1}x_1 + \cdots + a_{in}x_n \quad (i = 1, \ldots, m).$$

Estenderemos agora essas considerações a uma transformação linear entre dois quaisquer espaços vetoriais de dimensão finita.

Sejam E, F espaços vetoriais de dimensão finita e $A\colon E \to F$ uma transformação linear. Fixadas bases $\mathcal{V} = \{v_1, \ldots, v_n\} \subset E$ e $\mathcal{W} =$

$\{w_1, \dots, w_m\} \subset F$, para cada $j = 1, \dots, n$ o vetor Av_j se exprime como combinação linear dos vetores da base $\mathcal{W}$:

$$Av_j = a_{1j}w_1 + a_{2j}w_2 + \cdots + a_{mj}w_m = \sum_{i=1}^{m} a_{ij}w_i \,.$$

Assim, a transformação linear $A\colon E \to F$ juntamente com as bases $\mathcal{V} \subset E$ e $\mathcal{W} \subset F$ determinam uma matriz $\mathbf{a} = [a_{ij}] \in M(m \times n)$, chamada a *matriz de* A relativamente a essas bases (ou nas bases $\mathcal{V}, \mathcal{W}$).

Por definição, a j-ésima coluna da matriz $\mathbf{a}$ é formada pelas coordenadas de Av_j em relação à base $\mathcal{W}$.

Embora isso não seja mencionado explicitamente, convém salientar que os vetores nas bases $\mathcal{V}$ e $\mathcal{W}$ são dispostos numa ordem fixa, sem o que a matriz $\mathbf{a}$ não ficaria bem definida.

No caso em que $A\colon E \to E$ é um operador linear, a menos que seja feita menção explícita em contrário, considera-se apenas uma base $\mathcal{V} = \{v_1, \dots, v_n\} \subset E$ e a matriz $\mathbf{a} = [a_{ij}]$ do operador A relativamente à base $\mathcal{V}$ (ou na base $\mathcal{V}$) é definida pelas n igualdades

$$Av_j = \sum_{i=1}^{n} a_{ij}v_i \qquad (j = 1, \dots, n).$$

Neste caso, $\mathbf{a} \in M(n \times n)$ é a matriz quadrada $n \times n$ cuja j-ésima coluna é formada pelas coordenadas do vetor

$$Av_j = a_{1j}v_1 + a_{2j}v_2 + \cdots + a_{nj}v_n$$

na base $\mathcal{V}$.

Quando considerarmos uma transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ e dissermos apenas *a matriz de* A, estaremos significando a matriz de A relativamente às bases canônicas de $\mathbb{R}^n$ e $\mathbb{R}^m$. Caso utilizemos outras bases, isto será dito explicitamente.

Exemplo 8.1. Consideremos um espaço vetorial E, de dimensão finita. Dado $\alpha \in \mathbb{R}$, seja $A\colon E \to E$ o operador linear definido por $Av = \alpha v$ para todo $v \in E$. Relativamente a qualquer base $\mathcal{V} = \{v_1, \dots, v_n\} \subset E$, a matriz $\mathbf{a}$ do operador A é sempre a mesma, com números α na

diagonal e zeros fora dela:

$$\mathbf{a} = \begin{bmatrix} \alpha & 0 & \cdots & 0 \\ 0 & \alpha & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha \end{bmatrix}$$

O operador $A = \alpha I$ é o que se chama uma *homotetia* de razão α. Estes são os únicos operadores cujas matrizes independem da base dada. (Vide Exercício 8.35.)

Exemplo 8.2. Seja $P\colon E \to E$ a projeção sobre o subespaço F_1, paralelamente ao subespaço F_2. Sejam ainda $\mathcal{V}_1 \subset F_1$ e $\mathcal{V}_2 \subset F_2$ bases quaisquer desses subespaços. Então $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$ é uma base de E, relativamente à qual a matriz $\mathbf{p}$ de P tem os k primeiros termos da diagonal iguais a 1 ($k = \dim F_1$) e todos os demais termos (sobre a diagonal ou fora dela) iguais a zero. Analogamente, se $S\colon E \to E$ é a reflexão em torno de F_1 paralelamente a F_2, sua matriz $\mathbf{s}$ na base $\mathcal{V}$ tem os primeiros k termos da diagonal iguais a 1, os restantes iguais a -1 e todos os termos fora da diagonal iguais a zero.

A fixação das bases $\mathcal{V} \subset E$ e $\mathcal{W} \subset F$ determina portanto uma transformação

$$\varphi\colon \mathcal{L}(E;F) \to M(m \times n),$$

que faz corresponder a cada $A \in \mathcal{L}(E;F)$ sua matriz $\mathbf{a}$ nas bases $\mathcal{V}$, $\mathcal{W}$. A transformação φ é linear, ou seja, se $\mathbf{a}, \mathbf{b} \in M(m \times n)$ são as matrizes de $A, B \in \mathcal{L}(E;F)$ respectivamente e α, β são números reais, então a matriz de $A + B$ é $\mathbf{a} + \mathbf{b}$, a matriz de αA é $\alpha\mathbf{a}$ e, mais geralmente, a matriz de $\alpha A + \beta B$ é $\alpha\mathbf{a} + \beta\mathbf{b}$. Mais ainda, φ é um isomorfismo: a bijetividade de φ é assegurada pelo Teorema 4.1.

Convém observar que, no caso de $E = \mathbb{R}^n$ e $F = \mathbb{R}^m$, existe um par natural de bases (canônicas) nestes espaços, de modo que o isomorfismo $\varphi\colon \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m) \to M(m \times n)$ pode ser definido sem depender de escolhas arbitrárias. A cada transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ corresponde a matriz $\varphi(A) = [a_{ij}]$ cujo j-ésimo vetor-coluna é $A \cdot e_j = (a_{1j}, \ldots, a_{mj})$.

Em particular, a cada funcional linear $f\colon \mathbb{R}^n \to \mathbb{R}$ corresponde, de modo natural, uma matriz $[a_1, \ldots, a_n] \in M(1 \times n)$ ou, o que é o mesmo, um vetor $(a_1, \ldots, a_n)$. As correspondências entre a matriz $[a_1, \ldots, a_n]$ e o funcional f tal que $f(e_i) = a_i$, e entre f e o vetor

$(a_1, \ldots, a_n)$, são isomorfismos entre $M(1 \times n)$, $(\mathbb{R}^n)^*$ e $\mathbb{R}^n$, determinados pela base canônica de $\mathbb{R}^n$.

Entre transformações lineares, além das operações $A + B$ e αA, existe também a multiplicação BA. O isomorfismo φ faz corresponder ao produto BA o produto **ba** das matrizes de B e de A, segundo definiremos a seguir.

Sejam $u = (\alpha_1, \ldots, \alpha_n)$ e $v = (\beta_1, \ldots, \beta_n)$ vetores em $\mathbb{R}^n$. O *produto interno* de u por v é definido como o número

$$\langle u, v \rangle = \alpha_1\beta_1 + \cdots + \alpha_n\beta_n \,.$$

A noção geral de produto interno, suas propriedades e aplicações serão estudadas na Seção 10. Um caso particular será usado agora para introduzir o produto de duas matrizes.

Sejam $\mathbf{b} = [b_{ij}] \in M(m \times n)$ e $\mathbf{a} = [a_{ij}] \in M(n \times p)$ matrizes tais que o número de colunas de **b** é igual ao número de linhas de **a**. O *produto* da matriz **b** pela matriz **a** (nesta ordem) é a matriz $\mathbf{ba} = \mathbf{c} = [c_{ij}] \in M(m \times p)$, cujo ij-ésimo elemento

$$c_{ij} = b_{i1}a_{1j} + b_{i2}a_{2j} + \cdots + b_{in}a_{nj} = \sum_{k=1}^{n} b_{ik}a_{kj}$$

é o produto interno do i-ésimo vetor-linha de b pelo j-ésimo vetor-coluna de **a**.

Exemplo 8.3. Uma transformação linear $A \colon \mathbb{R}^n \to \mathbb{R}^m$ pode ser interpretada como uma multiplicação de matrizes: em vez de $A \in \mathcal{L}(\mathbb{R}^n; \mathbb{R}^m)$ considera-se sua matriz $\mathbf{a} = [a_{ij}] \in M(m \times n)$. Em particular, os funcionais lineares $f \colon \mathbb{R}^n \to \mathbb{R}$ são substituídos por matrizes $1 \times n$, ou seja, por vetores-linha. Além disso, os vetores $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ e $b = (b_1, \ldots, b_m)$ passam a ser considerados como matrizes $n \times 1$ e $m \times 1$ respectivamente, ou seja, como vetores-coluna. Então a igualdade $Ax = b$ passa a ser escrita sob a forma $\mathbf{ax} = \mathbf{b}$, isto é:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Dentro deste ponto de vista, a Álgebra Linear se reduz ao cálculo de matrizes, o que traz vantagens sob o aspecto computacional mas

o custo é a perda da intuição geométrica, da simplicidade conceitual, além da impossibilidade de se tratar o caso de dimensão infinita.

A definição do produto de matrizes foi formulada de modo a tornar verdadeiro o teorema seguinte. Nele, $A\colon E \to F$ e $B\colon F \to G$ são transformações lineares, $\mathcal{U} = \{u_1, \ldots, u_p\} \subset E$, $\mathcal{V} = \{v_1, \ldots, v_n\} \subset F$ e $\mathcal{W} = \{w_1, \ldots, w_m\} \subset G$ são bases, $\mathbf{a} \in M(n \times p)$ é a matriz de A nas bases $\mathcal{U}$, $\mathcal{V}$ e $\mathbf{b} \in M(m \times n)$ é a matriz de B nas bases $\mathcal{V}$, $\mathcal{W}$.

Teorema 8.1. *A matriz de* $BA\colon E \to G$ *nas bases* $\mathcal{U}$, $\mathcal{W}$ *é o produto* $\mathbf{ba} \in M(m \times p)$ *das matrizes* $\mathbf{b}$ *e* $\mathbf{a}$.

Demonstração: Por definição, temos

$$Au_j = \sum_{k=1}^{n} a_{kj} v_k \quad (j = 1, \ldots, p)$$

e

$$Bv_k = \sum_{i=1}^{m} b_{ik} w_i \quad (k = 1, \ldots, n).$$

Seja $\mathbf{c} = [c_{ij}] \in M(m \times p)$ a matriz de BA nas bases $\mathcal{U}$, $\mathcal{W}$. Por definição, para cada $j = 1, \ldots, p$, temos:

$$\begin{aligned} \sum_{i=1}^{m} c_{ij} w_i = BAu_j = B\left(\sum_{k=1}^{n} a_{kj} v_k\right) = \sum_{k=1}^{n} a_{kj} Bv_k = \\ = \sum_{k=1}^{m} a_{kj} \left(\sum_{i=1}^{n} b_{ik} w_i\right) = \sum_{i=1}^{n} \left[\sum_{k=1}^{m} b_{ik} a_{kj}\right] w_i\,. \end{aligned}$$

Igualando os coeficientes de cada w_i, concluímos que, para i=1, ..., m e $j = 1, \ldots, p$, tem-se

$$c_{ij} = \sum_{k=1}^{n} b_{ik} a_{kj}\,,$$

logo $\mathbf{c} = \mathbf{ba}$. □

Resulta imediatamente do teorema acima e do isomorfismo $\varphi\colon \mathcal{L}(E; F) \to M(m \times n)$ que as regras operacionais do produto de transformações lineares se transferem diretamente para o produto de matrizes. No que se segue, indicaremos com o símbolo $\mathbf{I}_n$ a *matriz*

identidade $n \times n$. Tem-se $\mathbf{I}_n = [\delta_{ij}]$, onde δ_{ij} é o *símbolo de Kronecker*: $\delta_{ij} = 0$ se $i \neq j$ e $\delta_{ii} = 1$. Quando não houver ambigüidade, escreveremos simplesmente $\mathbf{I}$ em vez de $\mathbf{I}_n$.

As propriedades abaixo listadas se provam considerando, para cada $\mathbf{a} \in M(m \times n)$, a transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ cuja matriz é $\mathbf{a}$ e aplicando a propriedade correspondente para transformações lineares, já provada anteriormente.

1) $(\mathbf{cb})\mathbf{a} = \mathbf{c}(\mathbf{ba})$;

2) $\mathbf{c}(\mathbf{a} + \mathbf{b}) = \mathbf{ca} + \mathbf{cb}$; $(\mathbf{b} + \mathbf{c})\mathbf{a} = \mathbf{ba} + \mathbf{ca}$;

3) $\mathbf{a} \cdot \mathbf{I}_n = \mathbf{a}$, $\mathbf{I}_m \cdot \mathbf{a} = \mathbf{a}$ *se* $\mathbf{a} \in M(m \times n)$;

4) $\mathbf{b}(\alpha\mathbf{a}) = \alpha(\mathbf{ba})$.

Dada $\mathbf{a} \in M(m \times n)$, diz-se que $\mathbf{x} \in M(n \times m)$ é uma matriz *inversa à esquerda* de $\mathbf{a}$ quando $\mathbf{xa} = \mathbf{I}_n$ e que $\mathbf{y} \in M(n \times m)$ é uma matriz *inversa à direita* de $\mathbf{a}$ quando $\mathbf{ay} = \mathbf{I}_m$.

5) *Uma matriz* $m \times n$ *possui inversa à esquerda se, e somente se, seus vetores-coluna são* L.I. *e uma inversa à direita se, e somente se, esses vetores-coluna geram* $\mathbb{R}^m$.

Uma matriz $\mathbf{a}$ chama-se *invertível* quando é quadrada e existe uma matriz $\mathbf{a}^{-1}$, chamada a *inversa* de $\mathbf{a}$, tal que $\mathbf{a}^{-1}\mathbf{a} = \mathbf{a}\mathbf{a}^{-1} = \mathbf{I}$.

6) *Se uma matriz* $\mathbf{a}$ *possui uma inversa à esquerda* $\mathbf{x}$ *e uma inversa à direita* $\mathbf{y}$ *então* $\mathbf{a}$ *é quadrada, é invertível e* $\mathbf{x} = \mathbf{y} = \mathbf{a}^{-1}$.

7) *Uma matriz quadrada* $\mathbf{a}$ *admite uma inversa à esquerda se, e somente se, admite uma inversa à direita. Neste caso, a matriz* $\mathbf{a}$ *é invertível e cada uma dessas inversas laterais é igual a* $\mathbf{a}^{-1}$.

A seguir, determinaremos como varia a matriz de uma transformação linear $A\colon E \to F$ quando se mudam as bases em E e F.

Sejam $\mathcal{V} = \{v_1, \ldots, v_n\} \subset E$ e $\mathcal{W} = \{w_1, \ldots, w_m\} \subset F$ bases, em relação às quais a matriz da transformação linear $A\colon E \to F$ é $\mathbf{a} = [a_{ij}] \in M(m \times n)$. Isto significa que

$$Av_j = \sum_{i=1}^{m} a_{ij} w_i \qquad (j = 1, \ldots, n).$$

Tomando novas bases $\mathcal{V}' = \{v'_1, \ldots, v'_n\} \subset E$ e $\mathcal{W}' = \{w'_1, \ldots, w'_m\} \subset F$, a transformação linear A tem nova matriz $\mathbf{a}' = [a'_{ij}] \in M(m \times n)$,

definida por:

$$Av'_j = \sum_{r=1}^{m} a'_{rj} w'_r \qquad (j = 1, \ldots, n). \tag{*}$$

Para obter a relação entre as matrizes $\mathbf{a}$ e $\mathbf{a}'$, consideramos as *matrizes de passagem* $\mathbf{p} = [p_{kj}] \in M(n \times n)$ e $\mathbf{q} = [q_{ir}] \in M(m \times m)$, definidas pelas igualdades

$$v'_j = \sum_{k=1}^{n} p_{kj} v_k \qquad \text{e} \qquad w'_r = \sum_{i=1}^{m} q_{ir} w_i \,.$$

Por definição, $\mathbf{p}$ é a matriz de passagem da base $\mathcal{V}$ para a base $\mathcal{V}'$ e $\mathbf{q}$ é a matriz de passagem da base $\mathcal{W}$ para a base $\mathcal{W}'$. Cada um dos dois membros da igualdade (*) pode ser escrito separadamente, em termos da base $\mathcal{W}$, assim:

$$\begin{aligned} Av'_j = \sum_{k=1}^{n} p_{kj}\, Av_k &= \sum_{k=1}^{n} p_{kj} \sum_{i=1}^{m} a_{ik} w_i \\ &= \sum_{k=1}^{n} \sum_{i=1}^{m} p_{kj} a_{ik}\, w_i \\ &= \sum_{i=1}^{m} \left(\sum_{k=1}^{n} a_{ik} p_{kj} \right) w_i \,, \end{aligned}$$

$$\begin{aligned} \sum_{r=1}^{m} a'_{rj} w'_r &= \sum_{r=1}^{m} a'_{rj} \sum_{i=1}^{m} q_{ir} w_i \\ &= \sum_{r=1}^{m} \sum_{i=1}^{m} a'_{rj} q_{ir}\, w_i \\ &= \sum_{i=1}^{m} \left(\sum_{r=1}^{m} q_{ir} a'_{rj} \right) w_i \,. \end{aligned}$$

Igualando os coeficientes de w_i vem:

$$\sum_{k=1}^{n} a_{ik} p_{kj} = \sum_{r=1}^{m} q_{ir} a'_{rj} \,,$$

isto é, $\mathbf{ap} = \mathbf{qa}'$.

Observemos agora que toda matriz de passagem é invertível. Com efeito, se $\mathbf{p}$ é a matriz de passagem da base $\mathcal{V}$ para a base $\mathcal{V}'$ então $\mathbf{p}$ é também a matriz (em relação à base $\mathcal{V}$) do operador linear $P\colon E \to E$, tal que $Pv_j = v'_j$ $(j = 1, \ldots, n)$, o qual é invertível porque leva uma base numa base.

Assim, da igualdade $\mathbf{ap} = \mathbf{qa}'$ podemos concluir

$$\mathbf{a}' = \mathbf{q}^{-1}\,\mathbf{ap}.$$

Esta é a fórmula que nos dá a matriz $\mathbf{a}'$ de A nas bases $\mathcal{V}'$, $\mathcal{W}'$ em função da matriz $\mathbf{a}$ de A nas bases $\mathcal{V}$, $\mathcal{W}$. No caso particular de um operador $A\colon E \to E$ e de suas matrizes $\mathbf{a}$, $\mathbf{a}'$ relativas às bases $\mathcal{V}$, $\mathcal{V}'$, temos uma única matriz de passagem $\mathbf{p}$, que nos dá

$$\mathbf{a}' = \mathbf{p}^{-1}\,\mathbf{ap}.$$

As duas matrizes quadradas $\mathbf{a}$ e $\mathbf{p}^{-1}\,\mathbf{ap}$ dizem-se *semelhantes*. Assim, as matrizes do mesmo operador em relação a bases diferentes são semelhantes. Vale também a recíproca: se $\mathbf{a}$ e $\mathbf{a}' = \mathbf{p}^{-1}\,\mathbf{ap}$ são matrizes $n \times n$ semelhantes então existe um operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$ tal que $\mathbf{a}$ e $\mathbf{a}'$ são matrizes de A relativamente a bases distintas de $\mathbb{R}^n$.

Com efeito, dadas $\mathbf{a}$ e $\mathbf{a}' = \mathbf{p}^{-1}\,\mathbf{ap}$, consideramos o operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$ cuja matriz na base canônica $\mathcal{E}$ de $\mathbb{R}^n$ é $\mathbf{a}$. Em seguida, consideramos a base $\mathcal{E}' \subset \mathbb{R}^n$, obtida da base canônica pela matriz de passagem $\mathbf{p}$. Então $\mathbf{a}'$ é a matriz de A na base $\mathcal{E}'$.

Para efeitos práticos é útil observar que se $\mathcal{V} = \{v_1, \ldots, v_n\}$ é uma base em $\mathbb{R}^n$ então a matriz de passagem da base canônica para $\mathcal{V}$ é aquela cujas n colunas são os vetores $v_1, \ldots, v_n$.

Exemplo 8.4. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador linear que consiste na reflexão em torno da reta $y = ax$. Como se viu no Exemplo 4.4, a matriz de A relativamente à base canônica de $\mathbb{R}^2$ é

$$\mathbf{a} = \begin{bmatrix} \frac{1-a^2}{1+a^2} & \frac{2a}{1+a^2} \\ \\ \frac{2a}{1+a^2} & \frac{a^2-1}{1+a^2} \end{bmatrix}.$$

Seja $\mathcal{V} = \{v_1, v_2\} \subset \mathbb{R}^2$ a base formada pelos vetores $v_1 = (1, a)$ e $v_2 = (-a, 1)$. Para todo vetor $v = (x, y) \in \mathbb{R}^2$, temos

$$A(x,y) = \left(\frac{1-a^2}{1+a^2}\,x + \frac{2a}{1+a^2}\,y, \frac{2a}{1+a^2}\,x + \frac{a^2-1}{1+a^2}\,y\right),$$

logo $Av_1 = v_1$ e $Av_2 = -v_2$. (De resto, estas igualdades são geometricamente óbvias.) Portanto a matriz de A na base $\mathcal{V}$ é

$$\mathbf{a}' = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

A matriz de passagem da base canônica de $\mathbb{R}^2$ para a base $\mathcal{V}$ é

$$\mathbf{p} = \begin{bmatrix} 1 & -a \\ a & 1 \end{bmatrix}.$$

Segue-se que $\mathbf{a}' = \mathbf{p}^{-1} \cdot \mathbf{a} \cdot \mathbf{p}$. Neste caso, foi mais simples calcular $\mathbf{a}'$ diretamente do que determinar $\mathbf{p}^{-1}$ e efetuar a multiplicação $\mathbf{p}^{-1}\mathbf{a}\mathbf{p}$. (Observação: $\mathbf{p}^{-1} = \frac{1}{1+a^2}\begin{bmatrix} 1 & a \\ -a & 1 \end{bmatrix}$.)

Seja $A\colon E \to F$ uma transformação linear entre espaços vetoriais de dimensão finita. O *posto* de A é a dimensão da sua imagem. Evidentemente, $\dim \mathcal{I}m(A) \leq \dim F$. Além disso, pelo Teorema do Núcleo e da Imagem, $\dim \mathcal{I}m(A) \leq \dim E$. Segue-se que o posto de A não excede $\dim E$ nem $\dim F$. O posto de A é igual à dimensão de E se, e somente se, A é injetiva. E é igual à dimensão de F se, e somente se, A é sobrejetiva.

Se $\mathbf{a} \in M(m \times n)$ é a matriz de A relativamente a um par qualquer de bases $\mathcal{U} \subset E$, $\mathcal{V} \subset F$, o posto de A é a dimensão do subespaço de $\mathbb{R}^m$ gerado pelas colunas de $\mathbf{a}$. Logo, o posto de A é o número máximo de colunas linearmente independentes da matriz $\mathbf{a}$.

Esta observação nos leva a definir o *posto segundo colunas* de uma matriz $\mathbf{a} \in M(m \times n)$ como o número máximo de colunas linearmente independentes em $\mathbf{a}$. Este número é igual à dimensão do subespaço vetorial de $\mathbb{R}^m$ gerado pelos vetores-coluna de $\mathbf{a}$. (*Espaço-coluna* de $\mathbf{a}$.)

De maneira análoga, definimos o *posto segundo linhas* da matriz $\mathbf{a} \in M(m \times n)$ como o número máximo de linhas L.I. em $\mathbf{a}$, ou seja, como a dimensão do subespaço vetorial de $\mathbb{R}^n$ gerado pelos vetores-linha da matriz $\mathbf{a}$. (*Espaço-linha* de $\mathbf{a}$.)

Embora o espaço-coluna e o espaço-linha da matriz $\mathbf{a}$ sejam subespaços de espaços vetoriais diferentes, vale o resultado seguinte:

Teorema 8.2. *Para toda matriz* $\mathbf{a} \in M(m \times n)$, *o posto segundo linhas e o posto segundo colunas são iguais.*

Demonstração: Seja p o posto segundo colunas da matriz $\mathbf{a} = [a_{ij}] \in M(m \times n)$. Então existem p vetores $w_k = (b_{1k}, \ldots, b_{mk}) \in \mathbb{R}^m$ tais que cada uma das colunas $v_j = (a_{1j}, \ldots, a_{mj})$, $1 \leq j \leq n$, é combinação linear de $w_1, \ldots, w_p$:

$$v_j = \sum_{k=1}^{p} c_{kj} w_k \,, \qquad 1 \leq j \leq n. \tag{*}$$

Tomando a i-ésima coordenada de cada um dos membros de (*), vemos que

$$a_{ij} = \sum_{k=1}^{p} c_{kj} b_{ik} = \sum_{k=1}^{p} b_{ik} c_{kj} \,, \tag{**}$$

para quaisquer i, j, com $1 \leq i \leq m$ e $1 \leq j \leq n$. Considerando agora os vetores-linha $u_i = (a_{i1}, \ldots, a_{in})$ da matriz $\mathbf{a}$, juntamente com os vetores $z_k = (c_{k1}, \ldots, c_{kn})$, $1 \leq k \leq p$, observamos que a igualdade entre o primeiro e o terceiro membro de (**) significa que, para todo $i = 1, \ldots, m$ tem-se

$$u_i = \sum_{k=1}^{p} b_{ik} z_k \,, \qquad 1 \leq i \leq m.$$

Assim, os vetores-linha de $\mathbf{a}$ são combinações lineares de $z_1, \ldots, z_p$, portanto o posto de $\mathbf{a}$ segundo linhas é $\leq p$. Aplicando este resultado à matriz $\mathbf{a}^T$ (chamada a *transposta* de $\mathbf{a}$), que tem como linhas as colunas de $\mathbf{a}$ e como colunas as linhas de $\mathbf{a}$, concluímos que o posto de $\mathbf{a}$ segundo colunas é menor do que ou igual ao posto segundo linhas. Isto conclui a demonstração. □

Podemos então definir o *posto* de uma matriz como o número máximo de linhas, ou de colunas, L.I. dessa matriz. Mesmo quando a matriz é quadrada (em cujo caso suas linhas e colunas pertencem ao mesmo espaço $\mathbb{R}^n$) os subespaços gerados pelas linhas e pelas colunas, respectivamente, podem ser diferentes mas têm a mesma dimensão.

Exemplo 8.5 (Uma aplicação do Teorema 8.2.) Sejam $f_1, ..., f_m : E \to \mathbb{R}$ funcionais lineares não-nulos no espaço vetorial E, de dimensão n. Vimos no Exemplo 6.10 que, para cada $i = 1, \ldots, m$, o núcleo de f_i é o subespaço vetorial, de dimensão $n - 1$, $H_i = \{v \in E; f_i(v) = 0\}$

(hiperplano em E). A interseção desses m hiperplanos é o subespaço vetorial $F = H_1 \cap \ldots \cap H_m$, formado pelos vetores $v \in E$ que cumprem simultaneamente as condições $f_1(v) = 0, \ldots, f_m(v) = 0$. Qual é a dimensão do subespaço F? Usando o Teorema 8.2 (juntamente com o Teorema do Núcleo e da Imagem), mostraremos agora que $\dim F = n - r$, onde r é o número máximo de elementos linearmente independentes no conjunto $\{f_1, \ldots, f_m\}$, isto é, a dimensão do subespaço de E^* gerado por estes funcionais.

Com efeito, fixemos uma base $\mathcal{V} = \{v_1, \ldots, v_n\} \subset E$ e seja $[a_{i1}, \ldots, a_{in}]$ a matriz de f_i nesta base ($i = 1, \ldots, m$). Temos $a_{ij} = f_i(v_j)$. Isto nos dá uma matriz $\mathbf{a} = [a_{ij}] \in M(m \times n)$, cuja i-ésima linha é a matriz de f_i, logo o posto de $\mathbf{a}$ segundo linhas é r. Segue-se do Teorema 8.2 que os vetores-coluna $w_1, \ldots, w_n$ de $\mathbf{a}$ geram um subespaço de dimensão r em $\mathbb{R}^m$. Ora, o subespaço gerado em $\mathbb{R}^m$ pelos w_j é a imagem da transformação linear $A \colon E \to \mathbb{R}^m$, definida por $Av = (f_1(v), \ldots, f_m(v))$, para todo $v \in E$. De fato,

$$
\begin{aligned}
Av_j &= (f_1(v_j), \ldots, f_m(v_j)) \\
&= (a_{1j}, \ldots, a_{mj}) = w_j, \quad j = 1, 2, \ldots, n.
\end{aligned}
$$

Evidentemente, o núcleo de A é o subespaço F. Resulta então do Teorema do Núcleo e da Imagem que $\dim F = \dim E - \dim \mathcal{I}m(A) = n - r$.

Exemplo 8.6. O espaço-linha e o espaço-coluna da matriz

$$
\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}
$$

são duas retas distintas em $\mathbb{R}^2$.

Exercícios

8.1. Determine a matriz do operador linear $A \colon \mathbb{R}^2 \to \mathbb{R}^2$, relativamente à base canônica, sabendo que $A(1,1) = (2,3)$ e $A(-1,1) = (4,5)$.

8.2. O produto vetorial de dois vetores $v = (x, y, z)$ e $w = (x', y', z')$ em $\mathbb{R}^3$ é, por definição, o vetor $v \times w = (yz' - zy',\ zx' - xz',\ xy' - yx')$.

Fixado o vetor $u = (a, b, c)$, determine a matriz, relativamente à base canônica, do operador $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, definido por $A \cdot v = v \times u$. Descreva geometricamente o núcleo desse operador e obtenha a equação da sua imagem.

8.3. Determine a matriz do operador de derivação $D\colon \mathcal{P}_n \to \mathcal{P}_n$ relativamente à base $\{1, t, t^2, \ldots, t^n\}$.

8.4. Considere os subespaços vetoriais F e G do espaço $C^\infty(\mathbb{R})$, cujas bases são, respectivamente, os conjuntos $\{\cos x, \operatorname{sen} x\}$ e

$$\{e^x \cos x, e^x \operatorname{sen} x, e^{2x} \cos x, e^{2x} \operatorname{sen} x, e^{3x} \cos x, e^{3x} \operatorname{sen} x\}.$$

Determine a matriz do operador de derivação em cada um desses subespaços.

8.5. Seja $A\colon E \to F$ uma transformação linear de posto r entre espaços vetoriais de dimensão finita. Prove que existem bases $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ e $\mathcal{V} = \{v_1, \ldots, v_m\} \subset F$, relativamente às quais a matriz $\mathbf{a} = [a_{ij}]$ de A tem $a_{11} = \cdots = a_{rr} = 1$ e os demais $a_{ij} = 0$.

8.6. Ache o valor de x para o qual operador $P\colon \mathbb{R}^3 \to \mathbb{R}^3$, cuja matriz na base canônica é

$$\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -1 & 0 & 1 \\ -\frac{1}{2} & -\frac{1}{2} & x \end{bmatrix}$$

seja uma projeção.

8.7. Qual é a matriz, na base canônica, do operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ tal que $A(2,3) = (2,3)$ e $A(-3,2) = 0$?

8.8. Calcule a n-ésima potência da matriz $\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$.

8.9. Seja $E = F_1 \oplus F_2$. Dado o operador linear $A\colon E \to E$, defina transformações lineares $A_{11}\colon F_1 \to F_1$, $A_{21}\colon F_1 \to F_2$, $A_{12}\colon F_2 \to F_1$ e $A_{22}\colon F_2 \to F_2$ tais que, para todo $v = v_1 + v_2 \in E$, com $v_1 \in F_1$ e $v_2 \in F_2$, seja

$$Av = (A_{11} + A_{21})v_1 + (A_{12} + A_{22})v_2 .$$

Diz-se então que

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

é a matriz do operador A relativamente à decomposição $E = F_1 \oplus F_2$. Dado outro operador linear $B \colon E \to E$, determine a matriz de BA relativamente à mesma decomposição.

8.10. Com a notação do exercício anterior, sejam **a** e **b** as matrizes de A e B respectivamente, em relação a uma base $\mathcal{U}_1 \cup \mathcal{U}_2 \subset E$, onde $\mathcal{U}_1 \subset F_1$ e $\mathcal{U}_2 \subset F_2$ são bases. Mostre que se

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{bmatrix} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_{11} & \mathbf{b}_{12} \\ \mathbf{b}_{21} & \mathbf{b}_{22} \end{bmatrix}$$

então

$$\mathbf{ba} = \begin{bmatrix} \mathbf{b}_{11}\mathbf{a}_{11} + \mathbf{b}_{12}\mathbf{a}_{21} & \mathbf{b}_{11}\mathbf{a}_{12} + \mathbf{b}_{12}\mathbf{a}_{22} \\ \mathbf{b}_{21}\mathbf{a}_{11} + \mathbf{b}_{22}\mathbf{a}_{21} & \mathbf{b}_{21}\mathbf{a}_{12} + \mathbf{b}_{22}\mathbf{a}_{22} \end{bmatrix}$$

(Multiplicação por blocos de matrizes.)

8.11. Seja **a** uma matriz 5×5 cujos elementos sobre a diagonal e abaixo dela são iguais a zero. Sem fazer nenhum cálculo, conclua que $\mathbf{a}^5 = 0$.

8.12. Sejam **a** uma matriz $m \times n$, com $m < n$, e **b** uma matriz $n \times m$. Podem **ab** e **ba** ser ambas invertíveis? Uma delas? Qual? Quando?

8.13. Assinale verdadeiro (V) ou falso (F):

( ) Se $A, B \colon E \to E$ são operadores de mesmo posto r então o produto BA tem posto r.

( ) Se as matrizes $\mathbf{a}, \mathbf{b} \in M(m \times n)$ têm o mesmo espaço-coluna então elas são matrizes da mesma transformação linear.

( ) A matriz do operador linear $A \colon E \to E$ na base $\{v_1, v_2, v_3, ..., v_n\}$ difere da matriz do mesmo operador na base $\{v_2, v_1, v_3, \ldots, v_n\}$ pela permutação das duas primeiras colunas.

( ) Sejam $\mathbf{a} \in M(2 \times 3)$ e $\mathbf{b} \in M(3 \times 2)$. Se $\mathbf{ab} = \mathbf{I}_2$ então $\mathbf{ba} = \mathbf{I}_3$.

( ) Se a matriz $\mathbf{a}'$ se obtém da matriz **a** por uma permutação de suas linhas então $\mathbf{a}'$ e **a** têm o mesmo posto.

8.14. Seguindo a orientação ali fornecida, prove as propriedades 1) a 7) da multiplicação de matrizes listadas após a demonstração do

Teorema 8.1. Em particular, prove que se $\mathbf{a}$ e $\mathbf{b}$ são matrizes $n \times n$ com $\mathbf{ba} = \mathbf{I}_n$ então se tem também $\mathbf{ab} = \mathbf{I}_n$. (Cfr. Corolário do Teorema 6.7.)

8.15. Sejam dados os vetores $v_1, \ldots, v_n \in \mathbb{R}^n$. Se, para cada $j = 1, \ldots, n$, o j-ésimo vetor da base canônica de $\mathbb{R}^n$ se exprime como

$$e_j = x_{1j}v_1 + \cdots + x_{nj}v_n\,,$$

prove que $\mathbf{x} = [x_{ij}]$ é a inversa da matriz que tem $v_1, \ldots, v_n$ como vetores-coluna.

8.16. Determine a inversa da matriz $\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{a} & \mathbf{I}_s \end{bmatrix}$ onde $\mathbf{a} \in M(s \times r)$ e $\mathbf{0} \in M(r \times s)$.

8.17. Sejam $\mathbf{a} \in M(m \times m)$ uma matriz de posto r e $\mathbf{b} \in M(n \times n)$ uma matriz de posto s. Prove que a matriz $(m+n) \times (m+n)$ abaixo tem posto $r + s$:

$$\begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{b} \end{bmatrix}$$

O símbolo $\mathbf{0}$ na primeira linha significa a matriz nula $m \times n$ e, na segunda linha, $\mathbf{0} \in M(n \times m)$.

8.18. Dadas $\mathbf{a} \in M(m \times m)$, $\mathbf{b} \in M(n \times n)$ e $\mathbf{c} \in M(m \times n)$, com posto de $\mathbf{a} = r$ e posto de $\mathbf{b} = s$, que postos pode ter a matriz abaixo?

$$\begin{bmatrix} \mathbf{a} & \mathbf{c} \\ \mathbf{0} & \mathbf{b} \end{bmatrix}$$

8.19. Seja $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, com $b \neq 0$, a matriz de um operador $A \colon \mathbb{R}^2 \to \mathbb{R}^2$ na base canônica. Ache uma base de $\mathbb{R}^2$ na qual a matriz de A seja $\begin{bmatrix} 0 & 1 \\ bc - ad & a + d \end{bmatrix}$.

8.20. Determine a matriz da projeção $P \colon \mathbb{R}^2 \to \mathbb{R}^2$, $P(x, y) = (x, 0)$ relativamente à base $\{u, v\} \subset \mathbb{R}^2$, onde $u = (1, 1)$ e $v = (1, 2)$.

8.21. Sabendo que a matriz do operador $A \colon \mathbb{R}^3 \to \mathbb{R}^3$ relativamente à base $\{u, v, w\} \subset \mathbb{R}^3$, onde $u = (1, 1, 1)$, $v = (1, 2, 1)$, $w = (1, 1, 3)$, é

$$\frac{1}{2}\begin{bmatrix} 3 & 1 & 3 \\ 0 & 2 & 0 \\ -1 & -1 & -1 \end{bmatrix}\,,$$

determine a matriz de A relativamente à base canônica de $\mathbb{R}^3$.

8.22. Obtenha bases $\mathcal{U} \subset \mathbb{R}^2$ e $\mathcal{V} \subset \mathbb{R}^3$ relativamente às quais a matriz da transformação linear $A\colon \mathbb{R}^2 \to \mathbb{R}^3$, dada por $A(x, y) = (2x + y, 3x - 2y, x + 3y)$, tem as linhas $(1, 0)$, $(0, 1)$ e $(0, 0)$.

8.23. Suponha que os operadores lineares $A, B\colon E \to E$ têm a mesma matriz $\mathbf{a} = [a_{ij}]$ em relação a duas bases $\mathcal{U}, \mathcal{V} \subset E$. Prove que existe um isomorfismo $C\colon E \to E$ tal que $B = CAC^{-1}$.

8.24. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador cuja matriz na base canônica é

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Prove que se

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

é a matriz de A relativamente a uma base qualquer de $\mathbb{R}^2$ então $a_{12} \neq 0$ ou $a_{21} \neq 0$. (Noutras palavras, nenhuma matriz de A é diagonal.)

8.25. Considere as transformações lineares

$$A\colon \mathbb{R}^{n+1} \to \mathcal{P}_n, \ \ A(\alpha_o, \alpha_1, \ldots, \alpha_n) = \alpha_o + \alpha_1 x + \cdots + \alpha_n x^n$$

$$B\colon \mathcal{P}_n \to \mathbb{R}^{n+1}, \ \ B.p(x) = (p(0), p(1), \ldots, p(n)).$$

Determine a matriz de $BA\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ (na base canônica) e prove que é uma matriz invertível.

8.26. Seja $\mathbf{a}$ a matriz $n \times n$ cujas linhas são os vetores $v_1 = (1, 2, \ldots$ $\ldots, n)$, $v_2 = (n + 1, n + 2, \ldots, 2n)$, etc. Prove que o posto de $\mathbf{a}$ é igual a 2 e que o subespaço de $\mathbb{R}^n$ gerado por suas linhas coincide com o subespaço gerado por suas colunas.

8.27. Prove que uma matriz $\mathbf{c} = [c_{ij}] \in M(m \times n)$ tem posto 1 se, e somente se, existem vetores não-nulos $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$ e $b = (b_1, \ldots, b_n) \in \mathbb{R}^n$ tais que $c_{ij} = a_i.b_j$ para todo i e todo j.

8.28. Assinale V(erdadeiro) ou F(also):

( ) Toda matriz é soma de matrizes de posto 1.

( ) O conjunto das matrizes de posto k em $M(m \times n)$ é um subespaço vetorial.

( ) A matriz

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{bmatrix}$$

tem posto 2 se, e somente se, existem i, j tais que $x_i y_j \neq x_j y_i$.

( ) Se $A \colon E \to E$ é um operador linear de posto 1 então $E = \mathcal{N}(A) \oplus \mathcal{I}m(A)$.

8.29. Prove que uma matriz $m \times n$ tem posto r se, e somente se, é possível selecionar r linhas e r colunas (porém não mais) de modo que os elementos comuns a elas formem uma matriz invertível $r \times r$. [Sugestão: reduza ao caso $r = \min\{m, n\}$ e aplique o Teorema 8.2.]

8.30. Sejam $f_1, \dots, f_m \colon E \to \mathbb{R}$ funcionais lineares no espaço vetorial E de dimensão n. Suponha que estes funcionais gerem em $E^* = \mathcal{L}(E; \mathbb{R})$ uma variedade afim de dimensão r. Prove que o conjunto F, formado pelos vetores $v \in E$ tais que

$$f_1(v) = f_2(v) = \cdots = f_m(v),$$

é um subespaço vetorial de dimensão $n - r + 1$.

8.31. Uma matriz $n \times n$ chama-se um *quadrado mágico* quando a soma dos elementos de cada uma de suas linhas, de cada coluna, da diagonal principal e da outra diagonal (ao todo $2n + 2$ somas) são iguais. Prove que, se $n \geq 3$, o conjunto Q_n dos quadrados mágicos $n \times n$ é um subespaço vetorial de dimensão $n^2 - 2n$ do espaço $M(n \times n)$. [Sugestão: use os Exercícios 8.30, 3.32 e 3.33.]

8.32. Em conformidade com o exercício anterior, determine os 8 elementos restantes da matriz 4×4 abaixo, de modo a obter um quadrado mágico

$$\begin{bmatrix} 1 & 2 & 3 & * \\ 4 & 5 & 6 & * \\ 7 & 8 & * & * \\ * & * & * & * \end{bmatrix}$$

8.33. Calcule o posto da matriz

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 2 & 1 & 0 \end{bmatrix}$$

e mostre que o subespaço gerado por suas linhas é diferente daquele gerado por suas colunas.

8.34. Obtenha números a, b, c tais que $ax + by + cz = 0$ seja a equação do plano gerado pelas colunas da matriz

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}.$$

8.35. Seja $A\colon E \to E$ um operador linear no espaço vetorial E, de dimensão finita. Supondo que A não seja um múltiplo do operador identidade, mostre que existem bases de E do tipo $\{v, Av, \ldots\}$ e $\{v, 2Av, \ldots\}$. Relativamente a estas bases, as matrizes de A são diferentes. Conclua que os operadores αI são os únicos cuja matriz não depende da base escolhida e que as matrizes do tipo $\alpha \mathbf{I}_n$ são as únicas que comutam com todas as matrizes invertíveis $n \times n$.

8.36. Seja $\mathbf{a}$ uma matriz triangular (isto é, $a_{ij} = 0$ se $i < j$) $n \times n$ cujos elementos da diagonal são todos iguais a zero. Mostre que $\mathbf{a}^n = 0$. [Sugestão: considere o operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$ cuja matriz na base canônica é $\mathbf{a}$.]

8.37. O *traço* de uma matriz quadrada $\mathbf{a} = [a_{ij}] \in M(n \times n)$ é a soma $\operatorname{tr} \mathbf{a} = a_{11} + \cdots + a_{nn}$ dos elementos da sua diagonal. Prove que $\operatorname{tr}(\mathbf{ab}) = \operatorname{tr}(\mathbf{ba})$ e conclua que todas as matrizes do mesmo operador $A\colon E \to E$, num espaço E de dimensão finita, têm o mesmo traço, o qual se indica com a notação $\operatorname{tr} A$.

8.38. Prove que o traço de um operador linear idempotente $P\colon E \to E$ é um número inteiro, igual ao seu posto.

8.39. Seja $\mathbf{c} = [c_{ij}] \in M(n \times n)$ uma matriz de posto 1. Prove que $\mathbf{c}^2 = (\operatorname{tr} \mathbf{c}).\mathbf{c}$ e, mais geralmente, para todo $n > 1$, $\mathbf{c}^n = (\operatorname{tr} \mathbf{c})^{n-1}\mathbf{c}$.

8.40. Sejam $\mathcal{U}$, $\mathcal{V}$ e $\mathcal{W}$ bases finitas do espaço vetorial E. Se $\mathbf{p}$ e $\mathbf{q}$ são respectivamente as matrizes de passagem de $\mathcal{U}$ para $\mathcal{V}$ e de $\mathcal{V}$ para $\mathcal{W}$, prove que as matrizes de passagem de $\mathcal{U}$ para $\mathcal{W}$ e de $\mathcal{V}$ para $\mathcal{U}$ são respectivamente $\mathbf{pq}$ e $\mathbf{p}^{-1}$.

8.41. Prove que o posto da transformação linear BA é menor do que ou igual ao posto de A e ao posto de B. Dê um exemplo em que posto de A = posto de B > posto de BA.

8.42. Dada a transformação linear $A\colon E \to F$, entre espaços de dimensão finita, sejam $E_1 \subset E$ e $F_1 \subset F$ subespaços tais que $E = \mathcal{N}(A) \oplus E_1$ e $F = \mathcal{I}m(A) \oplus F_1$. Tome bases $\mathcal{U} \subset E$ e $\mathcal{V} \subset F$ cujos primeiros elementos formam respectivamente uma base de $\mathcal{N}(A)$ e uma base de $\mathcal{I}m(A)$. Que forma tem a matriz de A relativamente a $\mathcal{U}$ e $\mathcal{V}$?

8.43. Sejam $A\colon E \to F$ e $B\colon F \to G$ transformações lineares entre espaços vetoriais de dimensão finita. Se B é injetiva, prove que o posto de BA é igual ao posto de A. Que condição sobre A assegura que o posto de BA seja igual ao de B?

8.44. Se E tem dimensão finita, prove que não existem operadores lineares $A, B\colon E \to E$ tais que $AB - BA = I$ ou tais que $AB - BA$ seja uma projeção. (Use o traço. Compare com o Exercício 5.13.)

8.45. Sejam $\mathcal{V}, \mathcal{V}' \subset E$, e $\mathcal{W}, \mathcal{W}' \subset F$ bases finitas, e $\mathbf{p}$, $\mathbf{q}$ as matrizes de passagem de $\mathcal{V}$ para $\mathcal{V}'$ e de $\mathcal{W}$ para $\mathcal{W}'$ respectivamente. Dada a transformação linear $A\colon E \to F$, sejam $\mathbf{a}$ e $\mathbf{a}'$ respectivamente as matrizes de A relativamente às bases $\mathcal{V}, \mathcal{W}$ e $\mathcal{V}', \mathcal{W}'$. Mostre que $\mathbf{p}$ é a matriz de I_E nas bases $\mathcal{V}', \mathcal{V}$ e $\mathbf{q}$ é a matriz de I_F nas bases $\mathcal{W}', \mathcal{W}$. Use as igualdades $A = AI_E$ e $A = I_F A$ para provar que $\mathbf{ap}$ e $\mathbf{qa}'$ são iguais à matriz de A nas bases $\mathcal{V}', \mathcal{W}$. Obtenha assim uma nova dedução da fórmula $\mathbf{a}' = \mathbf{q}^{-1}\mathbf{ap}$.

8.46. Prove que uma matriz quadrada de posto 1 é idempotente se, e somente se, seu traço é igual a 1 .

9

Eliminação

Esta seção trata de aspectos computacionais dos assuntos tratados até aqui. Do ponto de vista do encadeamento lógico, sua leitura não é necessária para o entendimento das seções seguintes. (Salvo no que tange à Seção 17 que, por sua vez, quase nada influi nas que lhe seguem.) Entretanto seu valor educativo é inestimável pois exibe um processo simples e bem sucedido para responder a perguntas naturais sobre subespaços, transformações lineares, sistemas de equações e matrizes.

Estudaremos a seguir algumas questões de natureza prática que serão resolvidas com o uso do tradicional e eficiente método de eliminação.

9.A. Dimensão do subespaço gerado por m vetores

A primeira questão que abordaremos é o problema de determinar a dimensão do subespaço gerado por m vetores $v_1, \ldots, v_m$ no espaço vetorial E que, por simplicidade, (porém sem perda de generalidade) suporemos ser o espaço euclidiano $\mathbb{R}^n$. Noutras palavras, queremos achar o número r tal que r dos vetores dados são linearmente independentes porém os demais são combinações lineares deles.

O princípio básico a ser utilizado é a observação óbvia de que se um dos vetores dados, digamos v_1, tem uma de suas coordena-

das, por exemplo a j-ésima, diferente de zero mas todos os demais vetores $v_2, \ldots, v_m$ têm a j-ésima coordenada nula então v_1 não é combinação linear de $v_2, \ldots, v_m$. Resulta então do Teorema 3.2 (ou melhor, da observação logo após) que se cada um dos vetores não-nulos $w_1, \ldots, w_r$ tem uma coordenada diferente de zero e a mesma coordenada é zero em todos os vetores seguintes a ele nesta lista então $\{w_1, \ldots, w_r\}$ é L.I..

Exemplo 9.1. Sejam $v_1 = (0, 1, 2, 3, 4)$, $v_2 = (0, 0, 1, 2, 3)$, e $v_3 = (0, 0, 0, 0, 1)$. Neste caso, a segunda coordenada de v_1 é 1 mas as segundas coordenadas de v_2 e v_3 são nulas. A terceira coordenada de v_2 é 1 mas a terceira coordenada de v_3 é zero. Logo $\{v_1, v_2, v_3\} \subset \mathbb{R}^5$ é um conjunto L.I..

O critério acima enunciado, que garante a independência linear dos vetores $w_1, \ldots, w_r \in \mathbb{R}^n$, pode ser refraseado assim: a primeira coordenada não-nula de cada w_i tem índice menor do que a primeira coordenada não-nula dos vetores subseqüentes $w_{i+1}, \ldots, w_r$.

Se, para cada $i = 1, \ldots, r$, escrevermos $w_i = (a_{i1}, \ldots, a_{in})$, teremos uma matriz $\mathbf{a} = [a_{ij}] \in M(r \times n)$, cujos r vetores-linha são $w_1, \ldots, w_r$. Diremos que essa matriz é *escalonada* quando o primeiro elemento não-nulo de cada uma de suas linhas está à esquerda do primeiro elemento não-nulo de cada uma das linhas subseqüentes e, além disso, as linhas nulas (se houver) estão abaixo das demais.

Com esta definição, podemos dizer que as linhas não-nulas de uma matriz escalonada são vetores linearmente independentes, ou seja, uma matriz escalonada $r \times n$ tem posto r se suas linhas forem todas diferentes de zero.

Exemplo 9.2. As matrizes abaixo são escalonadas:

$$\begin{bmatrix} 1 & 3 & 7 & 2 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 & 3 & 1 \\ 0 & 0 & 4 & 5 & 2 \\ 0 & 0 & 0 & 6 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Ambas têm posto 3.

Dados os vetores $v_1, \ldots, v_m \in \mathbb{R}^n$, vamos alterá-los passo a passo de tal modo que, em cada etapa, os vetores obtidos geram o mesmo subespaço que os da etapa anterior e, no final, os vetores resultantes formam as linhas de uma matriz escalonada. Os não-nulos dentre

eles formarão uma base do subespaço gerado pelos vetores originalmente dados.

As seguintes modificações, chamadas *operações elementares*, levam os vetores $v_1, \ldots, v_m \in \mathbb{R}^n$ em vetores $v'_1, \ldots, v'_m \in \mathbb{R}^n$ que geram o mesmo subespaço: $S(v'_1, \ldots, v'_m) = S(v_1, \ldots, v_m)$.

(1) Trocar a posição de dois vetores v_i, v_j $(i < j)$ na lista dada. Esta operação é esquematizada como

$$(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_m) \mapsto (v_1, \ldots, v_j, \ldots, v_i, \ldots, v_m).$$

(2) Somar a um dos vetores um múltiplo de outro vetor da lista, ou seja, substituir v_j por $v'_j = v_j + \alpha v_i$, $i \neq j$.

Para justificar a operação (2), sejam $\mathcal{V} = (v_1, \ldots, v_m)$ e $\mathcal{V}' = (v_1, \ldots, v'_j, \ldots, v_m)$. Evidentemente $S(\mathcal{V}') \subset S(\mathcal{V})$. Além disso, como $v_j = v'_j - \alpha v_i$, segue-se que $S(\mathcal{V}) \subset S(\mathcal{V}')$. Logo $\mathcal{V}$ e $\mathcal{V}'$ geram o mesmo subespaço: $S(\mathcal{V}) = S(\mathcal{V}')$.

Em termos da matriz cujas linhas são os vetores dados, estas operações elementares se exprimem assim:

(1) Trocar a posição de duas linhas;

(2) Somar a uma linha um múltiplo de outra linha.

Portanto, o subespaço gerado pelas linhas (ou seja, o espaço-linha) de uma matriz não se altera quando essas duas operações elementares são aplicadas a essa matriz.

Descreveremos a seguir o processo de *eliminação* (ou *escalonamento*), o qual, mediante aplicações sucessivas das duas operações elementares às linhas de uma matriz, produz uma matriz escalonada. O procedimento é o seguinte:

(a) Se $a_{11} \neq 0$, o processo começa deixando a primeira linha intacta e somando a cada linha L_i, com $i \geq 2$, a primeira linha multiplicada por $-a_{i1}/a_{11}$. Com isto se obtém uma matriz cuja primeira coluna é $(a_{11}, 0, \ldots, 0)$.

(b) Se $a_{11} = 0$, uma troca de linhas fornece uma matriz com $a_{11} \neq 0$, desde que a primeira coluna não seja nula. Se, porém, todos os elementos da primeira coluna são iguais a zero, passa-se para a segunda coluna ou, mais geralmente, para a coluna mais próxima, à direita da primeira, onde haja algum elemento não-nulo e opera-se

como antes, de modo a obter uma matriz cuja primeira coluna não-nula começa com elemento $\neq 0$ mas todos os demais são iguais a zero. A partir daí não se mexe mais na primeira linha. Recomeça-se o processo, trabalhando com as linhas a partir da segunda, até obter uma matriz escalonada.

Exemplo 9.3. Sejam os vetores

$$\begin{aligned} v_1 &= (1,2,3,4), \\ v_2 &= (5,6,7,8) \text{ e} \\ v_3 &= (9,10,11,12) \end{aligned}$$

em $\mathbb{R}^4$. Indicamos abaixo a seqüência de operações elementares efetuadas sobre a matriz cujas linhas são estes vetores, conduzindo a uma matriz escalonada

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \xrightarrow[L_3-9L_1]{L_2-5L_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & -8 & -16 & -24 \end{bmatrix} \xrightarrow{L_3-2L_2}$$

$$\xrightarrow{L_3-2L_2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Como a matriz escalonada final tem duas linhas diferentes de zero, os três vetores dados geram um subespaço vetorial de dimensão 2 em $\mathbb{R}^4$ e $w_1 = (1,2,3,4)$, $w_2 = (0,-4,-8,-12)$ formam uma base desse subespaço.

No exemplo acima, como nos seguintes, a notação $L_i + \alpha L_j$ significa que a matriz à direita foi obtida da matriz à esquerda somando-se à i-ésima linha o múltiplo αL_j da j-ésima linha. Analogamente, usaremos a notação $L_i \leftrightarrow L_j$ para indicar a troca da linha i pela linha j.

Exemplo 9.4. Consideremos os vetores $v_1 = (0,1,2,3)$, $v_2 = (2,1,3,0)$, $v_3 = (3,4,2,0)$ e $v_4 = (4,2,0,1)$ em $\mathbb{R}^4$. Indicamos abaixo a seqüência de operações elementares efetuadas sobre a matriz que

tem esses vetores como linhas a fim de obter uma matriz escalonada

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 2 & 1 & 3 & 0 \\ 3 & 4 & 2 & 0 \\ 4 & 2 & 0 & 1 \end{bmatrix} \xrightarrow{L_2 \leftrightarrow L_1} \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 3 & 4 & 2 & 0 \\ 4 & 2 & 0 & 1 \end{bmatrix} \xrightarrow[L_4 - 2L_1]{L_3 - \frac{3}{2}L_1} \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & \frac{5}{2} & -\frac{5}{2} & 0 \\ 0 & 0 & -6 & 1 \end{bmatrix}$$

$$\xrightarrow{L_3 - \frac{5}{2}L_2} \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} \\ 0 & 0 & -6 & 1 \end{bmatrix} \xrightarrow{L_4 - \frac{4}{5}L_3} \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} \\ 0 & 0 & 0 & 7 \end{bmatrix}.$$

Concluímos que os quatro vetores dados são L.I., portanto constituem uma base de $\mathbb{R}^4$. Além disso, vemos que os vetores $w_1 = (2,1,3,0)$, $w_2 = (0,1,2,3)$, $w_3 = (0,0,-\frac{15}{2},-\frac{15}{2})$ e $w_4 = (0,0,0,7)$ também formam uma base de $\mathbb{R}^4$.

9.B. Cálculo do posto de uma transformação linear

A resposta à questão 9.A permite determinar o posto de uma transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ e até mesmo uma base para $\mathcal{I}m(A)$. Uma tal base pode ser formada pelas colunas não-nulas de uma matriz escalonada, obtida da matriz de A por meio de operações elementares efetuadas sobre suas *colunas*. Ou então podemos, como acima, operar sobre as linhas da transposta da matriz de A. (Pois as linhas da transposta são as colunas da matriz dada.) Não haverá confusão se lembrarmos que a base de $\mathcal{I}m(A)$ é formada por vetores de $\mathbb{R}^m$, não de $\mathbb{R}^n$! Quando $m = n$, é preciso ter cuidado, pois a imagem de A é gerada pelos vetores-coluna de sua matriz e não pelos vetores-linha.

Exemplo 9.5. Obter uma base para a imagem da transformação linear $A\colon \mathbb{R}^3 \to \mathbb{R}^4$, definida por

$$A(x,y,z) = (x + 5y + 9z, 2x + 6y + 10z, 3x + 7y + 11z, 4x + 8y + 12z).$$

Temos $Ae_1 = (1,2,3,4)$, $Ae_2 = (5,6,7,8)$ e $Ae_3 = (9,10,11,12)$, de modo que a imagem de A é gerada pelos vetores v_1, v_2, v_3 do Exemplo 9.3. Resulta então daquele exemplo que A tem posto 2 e os vetores $w_1 = (1,2,3,4)$, $w_2 = (0,-4,-8,-12)$ formam uma base de $\mathcal{I}m(A)$. Note que a matriz que ocorre no Exemplo 9.3 não é a matriz de A e sim a sua transposta.

9.C. Resolução de sistemas lineares

O método de eliminação, embora simples e ingênuo, é a maneira mais eficaz de resolver um sistema de m equações lineares, com n incógnitas, apresentado sob a forma matricial $\mathbf{ax} = \mathbf{b}$, onde $\mathbf{a} \in M(m \times n)$, $\mathbf{x} \in M(n \times 1)$ e $\mathbf{b} \in M(m \times 1)$.

Resulta das noções gerais até aqui estudadas que o sistema $\mathbf{ax} = \mathbf{b}$ possui solução se, e somente se, o vetor $b \in \mathbb{R}^m$, correspondente à matriz $\mathbf{b}$, pertence à imagem da transformação linear $A\colon \mathbb{R}^n \to \mathbb{R}^m$ cuja matriz (nas bases canônicas de $\mathbb{R}^n$ e $\mathbb{R}^m$) é $\mathbf{a}$.

Dito de outra maneira, o sistema $\mathbf{ax} = \mathbf{b}$ possui solução se, e somente se, o vetor $b \in \mathbb{R}^m$ (correspondente à matriz $\mathbf{b}$) pertence ao subespaço gerado pelas colunas de $\mathbf{a}$. Isto equivale a dizer que a *matriz aumentada* $[\mathbf{a}; \mathbf{b}] \in M(m \times (n+1))$ tem o mesmo posto que a matriz $\mathbf{a}$ do sistema.

Uma afirmação mais completa é a seguinte: o sistema $Ax = b$ não possui solução quando $b \notin \mathcal{I}m(A)$, possui uma única solução quando $b \in \mathcal{I}m(A)$ e A é injetiva, e possui infinitas soluções quando $b \in \mathcal{I}m(A)$ e A não é injetiva. (Vide Teorema 6.4.)

Em termos matriciais, o sistema $\mathbf{ax} = \mathbf{b}$, com $\mathbf{a} \in M(m \times n)$, $\mathbf{x} \in M(n \times 1)$ e $\mathbf{b} \in M(m \times 1)$, admite as seguintes alternativas:

(1) Não possui solução quando o posto da matriz aumentada $[\mathbf{a}; \mathbf{b}]$ é maior do que o posto de $\mathbf{a}$;

(2) Possui uma única solução quando a matriz $\mathbf{a}$ e a matriz aumentada $[\mathbf{a}; \mathbf{b}]$ têm o mesmo posto, igual ao número n de incógnitas;

(3) possui infinitas soluções quando se tem posto $[\mathbf{a}; \mathbf{b}] =$ posto $\mathbf{a} = r < n$. Neste caso, o conjunto das soluções é uma variedade afim de dimensão $n - r$.

O que acabamos de dizer é mais ou menos um resumo do que já vimos antes. Trata-se de uma discussão esclarecedora do ponto de vista teórico mas que não ensina como reconhecer, na prática, em qual dos casos se enquadra um sistema dado e muito menos como obter suas soluções, caso existam. Isto se faz com o método de eliminação, escalonando a matriz aumentada do sistema.

O processo de eliminação se baseia na observação de que ao efetuar uma operação elementar sobre as linhas da matriz aumentada $[\mathbf{a}; \mathbf{b}]$ obtém-se uma matriz $[\mathbf{a}'; \mathbf{b}']$ que é a matriz aumentada de um sistema $\mathbf{a}'\mathbf{x} = \mathbf{b}'$, equivalente ao sistema original $\mathbf{ax} = \mathbf{b}$. (Dois sis-

temas se dizem *equivalentes* quando possuem o mesmo conjunto de soluções.)

No final do processo, obtém-se um sistema $\mathbf{a}'\mathbf{x} = \mathbf{b}'$, equivalente ao sistema proposto $\mathbf{ax} = \mathbf{b}$, no qual a matriz $[\mathbf{a}'; \mathbf{b}']$ é escalonada. (Isto é o mesmo que dizer que $\mathbf{a}'$ é escalonada.) O sistema $\mathbf{a}'\mathbf{x} = \mathbf{b}'$ é facilmente resolvido de baixo para cima: acha-se primeiro o valor da última incógnita, substituindo-a por esse valor na equação anterior e assim por diante.

Vejamos alguns exemplos.

Exemplo 9.6. Consideremos o sistema

$$\begin{array}{rcrcrcrcr} & & y & + & 2z & + & 3t & = & 1 \\ 2x & + & y & + & 3z & & & = & 1 \\ 3x & + & 4y & + & 2z & & & = & 1 \\ 4x & + & 2y & & & + & t & = & 1 \end{array}$$

O escalonamento da matriz aumentada é feito abaixo:

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 1 \\ 2 & 1 & 3 & 0 & 1 \\ 3 & 4 & 2 & 0 & 1 \\ 4 & 2 & 0 & 1 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 1 & 3 & 0 & 1 \\ 0 & 1 & 2 & 3 & 1 \\ 3 & 4 & 2 & 0 & 1 \\ 4 & 2 & 0 & 1 & 1 \end{bmatrix} \longrightarrow$$

$$\longrightarrow \begin{bmatrix} 2 & 1 & 3 & 0 & 1 \\ 0 & 1 & 2 & 3 & 1 \\ 0 & \frac{5}{2} & -\frac{5}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & -6 & 1 & -1 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 1 & 3 & 0 & 1 \\ 0 & 1 & 2 & 3 & 1 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} & -3 \\ 0 & 0 & 0 & 7 & \frac{7}{5} \end{bmatrix}.$$

Obtém-se assim a matriz aumentada do sistema

$$\begin{array}{rcrcrcrcr} 2x & + & y & + & 3z & & & = & 1 \\ & & y & + & 2z & + & 3t & = & 1 \\ & & & - & \frac{15}{2}z & - & \frac{15}{2}t & = & -3 \\ & & & & & & 7t & = & \frac{7}{5}. \end{array}$$

Resolvendo este sistema de baixo para cima, vem: $t = \frac{1}{5}$, $z = \frac{1}{5}$, $y = 0$, $x = \frac{1}{5}$. Esta é a única solução do sistema dado. Como a matriz do sistema tem posto 4, a solução existiria e seria única, fosse qual fosse o segundo membro.

Exemplo 9.7. Seja o sistema

$$\begin{array}{rcrcrcr} x & + & 2y & - & 3z & = & 4 \\ 2x & + & 3y & + & 4z & = & 5 \\ 4x & + & 7y & - & 2z & = & 12. \end{array}$$

O escalonamento da sua matriz aumentada é o seguinte:

$$\begin{bmatrix} 1 & 2 & -3 & 4 \\ 2 & 3 & 4 & 5 \\ 4 & 7 & -2 & 12 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -3 & 4 \\ 0 & -1 & 10 & -3 \\ 0 & -1 & 10 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -3 & 4 \\ 0 & -1 & 10 & -3 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Vemos portanto que o sistema dado é equivalente a:

$$\begin{array}{rcrcrcr} x & + & 2y & - & 3z & = & 4 \\ & - & y & + & 10z & = & -3 \\ 0x & + & 0y & + & 0z & = & -1, \end{array}$$

o qual é obviamente impossível. O sistema dado não tem solução. [Poderíamos ter chegado à mesma conclusão observando, na forma do Exemplo 9.5, que a imagem do operador $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, cuja matriz tem colunas $v_1 = (1,2,4)$, $v_2 = (2,3,7)$ e $v_3 = (-3,4,-2)$, é um subespaço de dimensão 2 em $\mathbb{R}^3$ do qual os vetores $w_1 = (1,2,4)$ e $w_2 = (0,1,1)$ formam uma base e que o vetor $b = (4,5,12)$ certamente não é combinação linear de w_1 e w_2.]

Exemplo 9.8. Seja o sistema

$$\begin{array}{rcrcrcrcr} x & + & 2y & + & 3z & + & 4t & = & 1 \\ 5x & + & 6y & + & 7z & + & 8t & = & 2 \\ 9x & + & 10y & + & 11z & + & 12t & = & 3 \end{array}$$

O escalonamento da sua matriz aumentada segue o esquema:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 1 \\ 5 & 6 & 7 & 8 & 2 \\ 9 & 10 & 11 & 12 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 3 & 4 & 1 \\ 0 & -4 & -8 & -12 & -3 \\ 0 & -8 & -16 & -24 & -6 \end{bmatrix} \longrightarrow$$

$$\longrightarrow \begin{bmatrix} 1 & 2 & 3 & 4 & 1 \\ 0 & -4 & -8 & -12 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A última matriz obtida é a matriz aumentada do sistema:

$$\begin{array}{rcrcrcrcr} x & + & 2y & + & 3z & + & 4t & = & 1 \\ & - & 4y & - & 8z & - & 12t & = & -3 \end{array}$$

ou

$$\begin{array}{rcrcrcrcr} x & + & 2y & = & -3z & - & 4t & + & 1 \\ & - & 4y & = & 8z & + & 12t & - & 3. \end{array}$$

Este sistema pode ser resolvido de baixo para cima (esquecendo que z e t são incógnitas) e nos dá a solução:

$$y = -2z - 3t + \frac{3}{4}, \qquad x = z + 2t - \frac{1}{2}. \qquad (*)$$

O sistema dado possui portanto uma infinidade de soluções, que podem ser obtidas atribuindo-se valores arbitrários a z e t e calculando x e y em função delas por meio destas duas últimas igualdades. Observe que as igualdades (*) são as equações da variedade afim de dimensão 2 no espaço $\mathbb{R}^4$, formada por todas as soluções do sistema dado. Escrevendo o sistema original sob a forma $Av = b$, onde $A\colon \mathbb{R}^4 \to \mathbb{R}^3$ é a transformação linear cuja matriz tem as linhas (1,2,3,4), (5,6,7,8), (9,10,11,12) e $b = (1,2,3)$, esta variedade afim, formada por todos os vetores

$$v = (z + 2t - \frac{1}{2}, -2z - 3t + \frac{3}{4}, z, t) \in \mathbb{R}^4,$$

onde z, t são números reais arbitrários, é o conjunto de todos os vetores $v \in \mathbb{R}^4$ tais que $Av = b$.

Observação: O conjunto $F = \{(z + 2t, -2z - 3t, z, t) \in \mathbb{R}^4; z, t \in \mathbb{R}\}$ é um subespaço vetorial de $\mathbb{R}^4$, núcleo da transformação linear $A\colon \mathbb{R}^4 \to \mathbb{R}^3$ acima considerada. Uma base de F é formada pelos vetores $w_1 = (1, -2, 1, 0)$ e $w_2 = (2, -3, 0, 1)$, obtidos fazendo $z = 1$, $t = 0$ e depois $z = 0$, $t = 1$ na expressão dos vetores de F. De um modo geral, para obter uma base para o núcleo de um operador $A\colon \mathbb{R}^n \to \mathbb{R}^m$ o que se tem a fazer é resolver por escalonamento o sistema $Ax = 0$.

Exemplo 9.9. Achar uma base para o núcleo da transformação linear $A\colon \mathbb{R}^5 \to \mathbb{R}^3$ cuja matriz (nas bases canônicas) é

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 3 & 4 & 5 & 3 & 4 \\ 1 & 0 & -1 & 1 & 0 \end{bmatrix}.$$

O núcleo de A é o conjunto das soluções $x = (x_1, x_2, x_3, x_4, x_5)$ do sistema linear homogêneo

$$\begin{array}{ccccccccccc} x_1 & + & 2x_2 & + & 3x_3 & + & x_4 & + & 2x_5 & = & 0 \\ 3x_1 & + & 4x_2 & + & 5x_3 & + & 3x_4 & + & 4x_5 & = & 0 \\ x_1 & & & - & x_3 & + & x_4 & & & = & 0. \end{array}$$

Para sistemas homogêneos, não há necessidade de considerar a matriz aumentada. O escalonamento da matriz **a** é feito segundo o esquema

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 3 & 4 & 5 & 3 & 4 \\ 1 & 0 & -1 & 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 0 & -2 & -4 & 0 & -2 \\ 0 & -2 & -4 & 0 & -2 \end{bmatrix} \longrightarrow$$

$$\longrightarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 0 & -2 & -4 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Portanto o sistema homogêneo inicial é equivalente ao sistema escalonado

$$\begin{array}{ccccccccccc} x_1 & + & 2x_2 & + & 3x_3 & + & x_4 & + & 2x_5 & = & 0 \\ & - & 2x_2 & - & 4x_3 & & & - & 2x_5 & = & 0, \end{array}$$

ou seja

$$\begin{array}{ccccccccc} x_1 & + & 2x_2 & = & -3x_3 & - & x_4 & - & 2x_5 \\ & - & 2x_2 & = & 4x_3 & & & + & 2x_5. \end{array}$$

Resolvendo o último (considerando x_3, x_4 e x_5 como conhecidos), vem $x_2 = -2x_3 - x_5$ e $x_1 = x_3 - x_4$. Concluímos então que o núcleo da transformação linear A é formado por todos os vetores $x = (x_3 - x_4, -2x_3 - x_5, x_3, x_4, x_5)$, onde os números x_3, x_4 e x_5 são escolhidos arbitrariamente. Uma base do núcleo é obtida quando se faz sucessivamente $(x_3, x_4, x_5) = (1, 0, 0)$, $(x_3, x_4, x_5) = (0, 1, 0)$ e $(x_3, x_4, x_5) = (0, 0, 1)$. Explicitamente, essa base é formada pelos vetores $w_1 = (1, -2, 1, 0, 0)$, $w_2 = (-1, 0, 0, 1, 0)$ e $w_3 = (0, -1, 0, 0, 1)$.

9.D. O método de Gauss-Jordan

A quarta aplicação que faremos do método de eliminação é o cálculo da inversa de uma matriz (invertível) $\mathbf{a} \in M(n \times n)$. Antes porém devemos advertir que a determinação da inversa não é necessária para resolver o sistema $\mathbf{ax} = \mathbf{b}$. A expressão $\mathbf{x} = \mathbf{a}^{-1} \cdot \mathbf{b}$ para a solução desse sistema é de grande elegância e significado teórico porém, na prática, a obtenção explícita da inversa $\mathbf{a}^{-1}$ requer a solução de n sistemas lineares. Convenhamos que isto seria um modo pouco eficaz de resolver um único sistema.

Com efeito, examinando coluna por coluna cada membro da igualdade $\mathbf{aa}^{-1} = \mathbf{I}_n$, vemos que a j-ésima coluna de $\mathbf{a}^{-1}$ é a solução do sistema $\mathbf{ax} = \mathbf{e}_j$, portanto o cálculo da inversa $\mathbf{a}^{-1}$ equivale a resolver os n sistemas lineares $\mathbf{ax} = \mathbf{e}_1, \ldots, \mathbf{ax} = \mathbf{e}_n$.

O método de eliminação que vimos utilizando é também chamado "método de Gauss". Existe ainda o "método de Gauss-Jordan".

Ele continua a eliminação iniciada pelo método de Gauss, chegando no final a uma matriz escalonada, com a propriedade adicional de que, acima e abaixo do primeiro elemento não-nulo de cada linha, todos os elementos são iguais a zero. Se a matriz for (quadrada e) invertível, o primeiro elemento não-nulo de cada linha da matriz escalonada está sobre a diagonal. Portanto, neste caso, o método de Gauss-Jordan produz uma matriz cujos elementos não-nulos constituem a diagonal.

Vejamos um exemplo da eliminação de Gauss-Jordan.

Exemplo 9.10. No Exemplo 9.4, o método de eliminação de Gauss em resumo operou a seguinte transformação por meio de operações elementares sobre as linhas:

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 2 & 1 & 3 & 0 \\ 3 & 4 & 2 & 0 \\ 4 & 2 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

O método de Gauss-Jordan continua, aplicando as operações elementares sobre as linhas, de modo a anular também os elementos de cada coluna situados acima da diagonal. Ele prossegue a partir

daí com as seguintes operações elementares:

$$\begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} \\ 0 & 0 & 0 & 7 \end{bmatrix} \xrightarrow{L_1 - L_2} \begin{bmatrix} 2 & 0 & 1 & -3 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} \\ 0 & 0 & 0 & 7 \end{bmatrix} \underset{L_2+\frac{4}{15}L_3}{\xrightarrow{L_1+\frac{2}{15}L_3}}$$

$$\underset{L_2+\frac{4}{15}L_3}{\xrightarrow{L_1+\frac{2}{15}L_3}} \begin{bmatrix} 2 & 0 & 0 & -4 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} \\ 0 & 0 & 0 & 7 \end{bmatrix} \underset{\substack{L_2-\frac{1}{7}L_4 \\ L_3+\frac{15}{14}L_4}}{\xrightarrow{L_1+\frac{4}{7}L_4}} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{15}{2} & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}.$$

Esta última matriz diagonal resulta portanto da matriz inicial pela aplicação sucessiva de operações elementares sobre suas linhas.

Existe ainda uma terceira operação elementar, que não tivemos ainda ocasião de mencionar porque não foi necessária até agora, mas que tem também a propriedade de, aplicada às linhas de uma matriz, não alterar o seu espaço-linha. Ela é a seguinte:

(3) Multiplicar uma linha por um número $\neq 0$.

Aplicando essa operação às linhas da matriz final do exemplo acima, obtemos a matriz identidade. (Multiplique a primeira linha por $1/2$, a terceira por $-2/15$ e a quarta por $1/7$.)

O método de Gauss-Jordan fornece imediatamente a solução do sistema $\mathbf{ax} = \mathbf{b}$ sem necessidade de, no final, efetuar a resolução de baixo para cima. Com efeito, depois de efetuada qualquer seqüência de operações elementares (inclusive a terceira) sobre as linhas da matriz aumentada obtemos sempre um sistema equivalente $\mathbf{a}'\mathbf{x} = \mathbf{b}'$. Se a matriz $\mathbf{a}$ é invertível, o processo de Gauss leva a uma matriz escalonada com elementos todos $\neq 0$ na diagonal. Prosseguindo a partir daí com Gauss-Jordan, chegaremos finalmente a um sistema $\mathbf{a}'\mathbf{x} = \mathbf{b}'$, equivalente ao original, com $\mathbf{a}' = \mathbf{I}_n$, logo $\mathbf{x} = \mathbf{b}'$, o que nos dá a solução $\mathbf{x}$ diretamente.

Assim, a solução do sistema $\mathbf{ax} = \mathbf{b}$ é a última coluna da matriz $[\mathbf{a}'; \mathbf{b}']$ que se obtém aplicando a eliminação de Gauss-Jordan à matriz aumentada $[\mathbf{a}; \mathbf{b}]$ de modo a chegar com $\mathbf{a}' = \mathbf{I}_n$.

Em particular, tomando $b = e_j$ (j-ésimo vetor da base canônica de $\mathbb{R}^n$), a solução x da equação $\mathbf{ax} = \mathbf{e}_j$, que é a j-ésima coluna de $\mathbf{a}^{-1}$, se obtém efetuando operações elementares sobre as linhas da matriz aumentada $[\mathbf{a}; \mathbf{e}_j]$ até reduzí-la a $[\mathbf{I}_n; \mathbf{x}]$. Como essas operações

dependem apenas da matriz **a** mas não de j, isto sugere o tópico seguinte

9.E. Método prático para calcular a inversa $\mathbf{a}^{-1}$

Acrescenta-se a matriz identidade $\mathbf{I}_n$ à direita de **a**, de modo a ter uma matriz aumentada $n \times 2n$:

$$\left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \dots & 1 \end{array}\right].$$

Em seguida aplicam-se operações elementares às linhas dessa matriz aumentada de modo a reduzir a matriz **a** à identidade $\mathbf{I}_n$, chegando-se a:

$$\left[\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & x_{11} & x_{12} & \dots & x_{1n} \\ 0 & 1 & \dots & 0 & x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & x_{n1} & x_{n2} & \dots & x_{nn} \end{array}\right].$$

A matriz $[x_{ij}]$ à direita é a inversa de **a**.

Exemplo 9.11. Damos abaixo um exemplo de como obter a inversa de uma matriz segundo este método.

$$\left[\begin{array}{ccc|ccc} 2 & 4 & 3 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 3 & 5 & 7 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 2 & 4 & 3 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & -1 & \frac{5}{2} & -\frac{3}{2} & 0 & 1 \end{array}\right] \to$$

$$\left[\begin{array}{ccc|ccc} 2 & 4 & 3 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} & -\frac{3}{2} & 1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 2 & 0 & 7 & 1 & -4 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} & -\frac{3}{2} & 1 & 1 \end{array}\right] \to$$

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 8 & -\frac{26}{3} & -\frac{14}{3} \\ 0 & 1 & 0 & -1 & \frac{5}{3} & \frac{2}{3} \\ 0 & 0 & \frac{3}{2} & -\frac{3}{2} & 1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & -\frac{13}{3} & -\frac{7}{3} \\ 0 & 1 & 0 & -1 & \frac{5}{3} & \frac{2}{3} \\ 0 & 0 & 1 & -1 & \frac{2}{3} & \frac{2}{3} \end{array}\right]$$

Portanto

$$\begin{bmatrix} 2 & 4 & 3 \\ 0 & 1 & -1 \\ 3 & 5 & 7 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & -\frac{13}{3} & -\frac{7}{3} \\ -1 & \frac{5}{3} & \frac{2}{3} \\ -1 & \frac{2}{3} & \frac{2}{3} \end{bmatrix}.$$

Retornaremos ao asssunto de eliminação gaussiana no item final da seção 17.

Exercícios

9.1. Determine o posto da matriz

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 17 \end{bmatrix}.$$

9.2. Ache a matriz, na base $\{u_1, u_2, u_3, u_4\}$, da projeção $P\colon \mathbb{R}^4 \to \mathbb{R}^4$, $P(x, y, z, t) = (x, y, 0, 0)$, sabendo que $u_1 = (2, 0, 3, 4)$, $u_2 = (1, 1, 4, 2)$, $u_3 = (3, 2, 2, 0)$ e $u_4 = (0, 3, 0, 1)$. [Sugestão: use eliminação gaussiana para exprimir e_1, e_2, e_3 e e_4 como combinações lineares de u_1, u_2, u_3 e u_4.]

9.3. Exprima cada um dos vetores da base canônica de $\mathbb{R}^3$ como combinação linear dos vetores $v_1 = (1, 1, 0)$, $v_2 = (-1, 0, 2)$, $v_3 = (4, 2, -5)$ e, a partir daí, obtenha a inversa da matriz $\begin{bmatrix} 1 & -1 & 4 \\ 1 & 0 & 2 \\ 0 & -2 & 5 \end{bmatrix}$.

9.4. Decida se as matrizes abaixo são invertíveis ou não. No caso afirmativo, determine a(s) inversa(s). Caso uma delas (digamos **a**) não seja invertível, ache uma matriz $\mathbf{x} \in M(3 \times 1)$ tal que $\mathbf{ax} = 0$:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 9 \\ 1 & 3 & 4 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 3 & 4 \end{bmatrix}.$$

9.5. Calcule a dimensão do subespaço vetorial de $\mathbb{R}^5$ gerado pelos vetores $v_1 = (2, 4, 8, -4, 7)$, $v_2 = (4, -2, -1, 3, 1)$, $v_3 = (3, 5, 2, -2, 4)$ e $v_4 = (-5, 1, 7, -6, 2)$. Decida se o vetor $b = (6, 18, 1, -9, 8)$ pertence ou não a esse subespaço.

9.6. A matriz $\mathbf{a} \in M(m \times n)$ tem apenas uma linha e uma coluna não-nulas. Dada $\mathbf{b} \in M(m \times 1)$, quais são as dimensões possíveis para a variedade afim formada pelas soluções $\mathbf{x} \in M(n \times 1)$ do sistema $\mathbf{ax} = \mathbf{b}$?

9.7. Exprima cada vetor do conjunto $\{u, v, w, z\} \subset E$ como combinação linear dos vetores $\{w, u + 3z, v - 2u + 3w, 5z\}$.

9.8. Obtenha uma base para o subespaço vetorial gerado por cada um dos seguintes conjuntos e, conseqüentemente, determine a dimensão daquele subespaço:

(a) $\{(1, 2, 3, 4), (3, 4, 7, 10), (2, 1, 3, 5)\}$

(b) $x^3 + 2x^2 + 3x + 4$, $5x^3 + 4x^2 + 3x + 2$, $4x^3 - 2x^2 + x$, $7x^3 + 2x^2 - 3x - 8$

(c) $(1, 3, 5)$, $(-1, 3, -1)$, $(1, 21, 1)$

(d) $(1, 2, 3)$, $(1, 4, 9)$, $(1, 8, 27)$.

9.9. Mostre que se 0, 1, a, b, c são números dois a dois diferentes então os vetores $(1, 1, 1, 1)$, (a, a^2, a^3, a^4), (b, b^2, b^3, b^4) e (c, c^2, c^3, c^4) são linearmente independentes. Generalize.

9.10. Exiba uma base para a imagem de cada uma das transformações lineares abaixo e determine seu posto.

(a) $A\colon \mathbb{R}^4 \to \mathbb{R}^3$, $A(x, y, z, t) = (x + 2y - t, 2x - z + 2t, -2x + y + 3z)$

(b) $B\colon \mathbb{R}^4 \to \mathbb{R}^5$, $B(x, y, z, t) = (x + 2y + 2z - t, 2x + 4y + 3z + t, -3x + 2z + 3t, -3x + z + 6t, 10x + 2y + 5z + 5t)$

(c) $C\colon \mathbb{R}^3 \to \mathbb{R}^3$, $C(x, y, z) = (x + 3y, 2y + 4z, x + y - 4z)$

(d) $D\colon \mathcal{P}_n \to \mathcal{P}_n$, $Dp(x) = p'(x)$.

9.11. Use escalonamento para resolver os seguintes sistemas linea-

res:

$$\begin{array}{rcrcrcl} x & + & 3y & + & z & = & 1 \\ 2x & + & 6y & + & 9z & = & 7 \\ 2x & + & 8y & + & 8z & = & 6 \end{array}$$

$$\begin{array}{rcrcrcrcr} x & + & y & & & + & t & = & 0 \\ x & + & 2y & + & z & + & t & = & 1 \\ 3x & + & 3y & + & z & + & 2t & = & -1 \\ & & y & + & 3z & - & t & = & 3 \end{array}$$

$$\begin{array}{rcrcrcrcl} x & + & y & - & z & + & 2t & = & 0 \\ & & 3y & - & z & + & 3t & = & 0 \\ 2x & - & y & - & z & + & t & = & 0 \end{array}$$

9.12. Ache uma condição envolvendo a, b, c para que o sistema abaixo tenha solução e encontre as soluções no caso em que elas existam

$$\begin{array}{rcrcrcrcl} x & + & y & + & z & + & t & = & a \\ & & 5y & + & 2z & + & 4t & = & b \\ 3x & - & 2y & + & z & - & t & = & c \end{array}$$

9.13. Ache uma base para o núcleo de cada uma das transformações lineares a seguir:

(a) $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, $A(x,y,z) = (-3y+4z, 3x-5z, -4x+5y)$.

(b) $B\colon \mathbb{R}^4 \to \mathbb{R}^5$, $B(x,y,z,t) = (2x-2z+4t, x-2z+3t, 4y+2z+t, 6x+4y-4z+13t, 2x+4y-2z+7t)$

(c) $C\colon \mathbb{R}^4 \to \mathbb{R}^3$, $C(x,y,z,t) = (2x+y-z+3t, x-4y+2z+t, 2y+4z-t)$.

(d) $T\colon \mathcal{P} \to \mathcal{P}$, $T \cdot p(x) = p(x+m)$, $m \neq 0$.

9.14. Decida quais das matrizes abaixo possuem inversa e calcule a inversa quando existir.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \begin{bmatrix} 4 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 8 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 4 & 3 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 2 & 1 \\ 3 & 1 & 1 & 2 \\ 1 & 2 & 1 & 3 \end{bmatrix}$$

9.15. Prove que o sistema

$$\begin{array}{rcrcrcrcl} x & + & 2y & + & 3z & - & 3t & = & a \\ 2x & - & 5y & - & 3z & + & 12t & = & b \\ 7x & + & y & + & 8z & + & 5t & = & c \end{array}$$

admite solução se, e somente se, $37a+13b = 9c$. Ache a solução geral do sistema quando $a = 2$ e $b = 4$.

9.16. Prove que toda matriz anti-simétrica 3×3 não-nula tem posto igual a dois. Dê exemplo de uma matriz anti-simétrica invertível 4×4.

9.17. Considere o sistema de n equações lineares a n incógnitas:

$$x_i + x_{i+1} = a_i \quad (i = 1, \ldots, n-1), \quad x_1 + x_n = a_n.$$

(a) Se n é ímpar, prove que ele possui solução única, sejam quais forem os a_i. Explicite esta solução.

(b) Supondo n par, obtenha condições sobre os a_i que sejam necessárias e suficientes para que o sistema possua solução. Caso existam soluções, determine a variedade afim por elas formada.

10

Produto Interno

O produto interno, que já foi mencionado brevemente antes, na definição do produto de duas matrizes, será apresentado formalmente nesta seção e adotado sistematicamente a partir daqui. Trata-se de uma noção que completa e enriquece a estrutura de um espaço vetorial, permitindo a utilização de uma linguagem geométrica altamente sugestiva e o destaque de tipos especiais de operadores, os quais admitem uma análise mais profunda de suas propriedades, como se verá a seguir.

Os axiomas de espaço vetorial não são suficientes para abordar certas noções geométricas como ângulo, perpendicularismo, comprimento, distância, etc. Isto se torna possível com a introdução de um produto interno.

Um *produto interno* num espaço vetorial E é um funcional bilinear simétrico e positivo em E. Mais precisamente, um produto interno é uma função $E \times E \rightarrow \mathbb{R}$, que associa a cada par de vetores $u, v \in E$ um número real $\langle u, v \rangle$, chamado o produto interno de u por v, de modo que sejam válidas as seguintes propriedades, para quaisquer $u, u', v, v' \in E$ e $\alpha \in \mathbb{R}$:

Bilinearidade: $\langle u + u', v \rangle = \langle u, v \rangle + \langle u', v \rangle$, $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$, $\langle u, v + v' \rangle = \langle u, v \rangle + \langle u, v' \rangle$, $\langle u, \alpha v \rangle = \alpha \langle u, v \rangle$;

Comutatividade (simetria): $\langle u, v \rangle = \langle v, u \rangle$;

Positividade: $\langle u, u\rangle > 0$ se $u \neq 0$.

Como $\langle 0, v\rangle = \langle 0+0, v\rangle = \langle 0, v\rangle + \langle 0, v\rangle$, segue-se que $\langle 0, v\rangle = \langle v, 0\rangle = 0$ para todo $v \in E$.

Resulta da positividade que se $\langle u, v\rangle = 0$ para todo $v \in E$ então $u = 0$. Com efeito, se fosse $u \neq 0$ teríamos $\langle u, v\rangle \neq 0$ pelo menos quando $v = u$.

Segue-se desta observação que se $u, u' \in E$ são vetores tais que $\langle u, v\rangle = \langle u', v\rangle$ para todo $v \in E$ então $u = u'$. Com efeito, isto implica que $\langle u - u', v\rangle = 0$ para todo $v \in E$, logo $u - u' = 0$ e $u = u'$.

O número não-negativo $|u| = \sqrt{\langle u, u\rangle}$ chama-se a *norma* ou o *comprimento* do vetor u. Com esta notação, tem-se $|u|^2 = \langle u, u\rangle$ e a igualdade

$$\langle u+v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle$$

lê-se: $|u+v|^2 = |u|^2 + |v|^2 + 2\,\langle u, v\rangle$.

Quando $|u| = 1$ diz-se que $u \in E$ é um *vetor unitário*. Todo vetor $u \neq 0$ se escreve como $u = |u| \cdot u'$, onde u' é um vetor unitário. Basta pôr $u' = |u|^{-1} \cdot u$.

Exemplo 10.1. No espaço euclidiano $\mathbb{R}^n$, o *produto interno canônico* dos vetores $u = (\alpha_1, \ldots, \alpha_n)$ e $v = (\beta_1, \ldots, \beta_n)$ é definido por $\langle u, v\rangle = \alpha_1\beta_1 + \cdots + \alpha_n\beta_n$. Este é o produto interno que consideraremos em $\mathbb{R}^n$, salvo aviso em contrário.

Exemplo 10.2. Consideremos $\mathbb{R}^2$ como o modelo aritmético do plano euclidiano, no qual se introduziu um sistema de coordenadas cartesianas. Dados $u = (\alpha_1, \alpha_2)$ e $v = (\beta_1, \beta_2)$, os números

$$|u| = \sqrt{\alpha_1^2 + \alpha_2^2}$$

e

$$|v| = \sqrt{\beta_1^2 + \beta_2^2}$$

medem realmente os comprimentos das flechas que representam esses vetores. Suponhamos $u \neq 0$, $v \neq 0$ e chamemos de θ o ângulo formado por essas flechas. Afirmamos que o produto interno $\langle u, v\rangle = \alpha_1\beta_1 + \alpha_2\beta_2$ acima definido é igual a $|u|\,|v| \cos\theta$. Isto será provado em três passos: 1º) Se os vetores u e v são perpendiculares, então $\langle u, v\rangle = 0 = |u|\,|v| \cos 90°$. Com efeito, por um lado,

$$|u+v|^2 = \langle u+v, u+v\rangle = |u|^2 + |v|^2 + 2\,\langle u, v\rangle$$

e por outro lado, pelo Teorema de Pitágoras,

$$|u + v|^2 = |u|^2 + |v|^2.$$

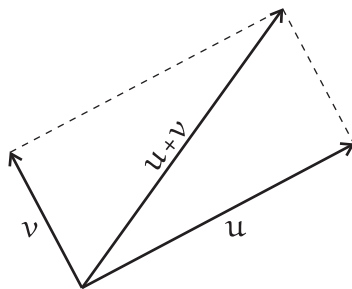



Figura 10.1.

Logo $\langle u, v\rangle = 0$. 2º) Se $|u| = |v| = 1$ então $\langle u, v\rangle = \cos\theta$. Com efeito, tomando o vetor unitário u^* perpendicular a u temos, pela definição de seno e cosseno, $v = \cos\theta \cdot u + \operatorname{sen}\theta \cdot u^*$. (Fig. 10.2.)

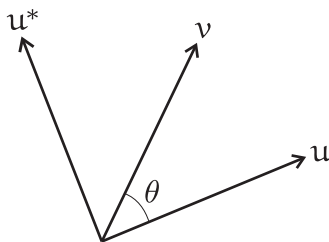



Figura 10.2.

Tomando o produto interno de ambos os membros desta igualdade por u vem $\langle u, v\rangle = \cos\theta \cdot \langle u, u\rangle + \operatorname{sen}\theta \cdot \langle u, u^*\rangle$. Como $\langle u, u\rangle = 1$ e $\langle u, u^*\rangle = 0$ pelo primeiro passo, temos $\langle u, v\rangle = \cos\theta$. 3º) Caso geral: pomos $u = |u| \cdot u'$ e $v = |v| \cdot v'$, onde $u' = (1/|u|)u$ e $v' = (1/|v|)v$ são vetores unitários. Então $\langle u, v\rangle = |u|\,|v|\,\langle u', v'\rangle = |u|\,|v| \cos\theta$. Vemos, em particular, que os vetores u, v formam um ângulo agudo quando $\langle u, v\rangle > 0$, um ângulo obtuso quando $\langle u, v\rangle < 0$ e um ângulo reto quando $\langle u, v\rangle = 0$.

Exemplo 10.3. Seja $E = C^o([a, b])$ o espaço vetorial cujos elementos são as funções contínuas $g, f\colon [a, b] \to \mathbb{R}$. Um produto interno em E pode ser definido pondo

$$\langle f, g \rangle = \int_a^b f(x)g(x)\, dx.$$

Neste caso, a norma da função f é

$$|f| = \sqrt{\int_a^b f(x)^2\, dx}.$$

Este produto interno é utilizado no estudo das séries de Fourier.

Observação. Seja E um espaço vetorial de dimensão finita arbitrário. Dada uma base $\{u_1, \ldots, u_n\} \subset E$, podemos definir um produto interno em E pondo, para $u = \Sigma\alpha_i u_i$ e $v = \Sigma\beta_i u_i$, $\langle u, v\rangle = \Sigma\alpha_i\beta_i$, por definição. Isto mostra que todo espaço vetorial de dimensão finita pode ser munido de um produto interno. (Fato verdadeiro em geral, pois qualquer espaço vetorial possui base, mas não entraremos nesse terreno.) Assim, quando nos referirmos a um espaço munido de um produto interno, não estaremos com isso atribuindo uma propriedade especial a esse espaço mas apenas dizendo que, entre os possíveis produtos internos que nele podem ser introduzidos, um particular foi escolhido e fixado.

Seja E um espaço vetorial com produto interno. Dois vetores $u, v \in E$ chamam-se *ortogonais* (ou *perpendiculares*) quando $\langle u, v\rangle = 0$. Escreve-se, então, $u \perp v$. Em particular, 0 é ortogonal a qualquer vetor de E. Um conjunto $X \subset E$ diz-se *ortogonal* quando dois vetores distintos quaisquer em X são ortogonais. Se, além disso, todos os vetores de X são unitários então X chama-se um conjunto *ortonormal*. Portanto, o conjunto $X \subset E$ é ortonormal se, e somente se, dados $u, v \in X$ tem-se $\langle u, v\rangle = 0$ se $u \neq v$ e $\langle u, v\rangle = 1$ se $v = u$. Uma *base ortonormal* é uma base de E que é um conjunto ortonormal.

Teorema 10.1. *Num espaço vetorial E com produto interno, todo conjunto ortogonal X de vetores não-nulos é* L.I..

Demonstração: Sejam $v_1, \ldots, v_n \in X$. Temos $\langle v_i, v_j \rangle = 0$ se $i \neq j$. Se $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ é uma combinação linear nula desses vetores então, para cada $i = 1, 2, \ldots, n$, tomamos o produto interno de ambos os membros desta igualdade por v_i e temos

$$\alpha_1 \langle v_1, v_i \rangle + \cdots + \alpha_n \langle v_n, v_i \rangle = 0,$$

logo $\alpha_i \langle v_i, v_i \rangle = \alpha_i |v_i|^2 = 0$ pois todos os produtos internos $\langle v_j, v_i \rangle$, com $j \neq i$, são nulos em virtude da ortogonalidade de X. Além disso, como os vetores pertencentes ao conjunto X são todos não-nulos, resulta de $\alpha_i |v_i|^2 = 0$ que $\alpha_i = 0$. Assim, os coeficientes da combinação linear $\Sigma \alpha_i v_i = 0$ são todos iguais a zero e os vetores do conjunto X são, portanto, linearmente independentes. □

Exemplo 10.4. A base canônica $\{e_1, \ldots, e_n\} \subset \mathbb{R}^n$ é ortonormal: tem-se $\langle e_i, e_j \rangle = \delta_{ij}$, onde $\delta_{ij} = 0$ se $i \neq j$ e $\delta_{ij} = 1$ se $i = j$. No plano $\mathbb{R}^2$ os vetores $u = (1, 1)$ e $v = (-1, 1)$ são ortogonais. Pondo

$$u' = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right) \quad \text{e} \quad v' = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right),$$

o conjunto $\{u', v'\} \subset \mathbb{R}^2$ é uma base ortonormal.

Quando u e v são ortogonais, a igualdade $|u + v|^2 = |u|^2 + |v|^2 + 2 \langle u, v \rangle$ se torna $|u + v|^2 = |u|^2 + |v|^2$. Esta é a versão do *Teorema de Pitágoras* para um espaço vetorial com produto interno.

Num espaço vetorial E com produto interno, seja u um vetor unitário. Dado qualquer $v \in E$, o vetor $\langle u, v \rangle \cdot u$ chama-se a *projeção ortogonal de* v *sobre o eixo que contém* u. A justificativa para esta denominação está no fato de que, escrevendo $w = v - \langle u, v \rangle u$, tem-se $v = \langle u, v \rangle u + w$, onde w *é perpendicular a* u. Com efeito, tomando o produto interno de u por ambos os membros da igualdade $w = v - \langle u, v \rangle u$ tem-se

$$\langle u, w \rangle = \langle u, v \rangle - \langle u, v \rangle \langle u, u \rangle = \langle u, v \rangle - \langle u, v \rangle = 0,$$

pois $\langle u, u \rangle = 1$. (Fig. 10.3)

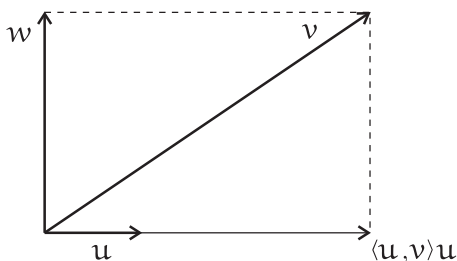



Figura 10.3.

Quando se tem apenas $u \neq 0$, o eixo que contém u é o mesmo que contém o vetor unitário $u' = u/|u|$ $(= |u|^{-1} \cdot u)$. A projeção ortogonal de v sobre este eixo é, portanto, igual a $\langle u', v\rangle\, u'$, ou seja, $(\langle u,v\rangle \,/\, \langle u,u\rangle) \cdot u$. Usaremos a notação

$$\text{pr}_u(v) = \frac{\langle u,v\rangle}{\langle u,u\rangle} \cdot u$$

para indicar a projeção ortogonal do vetor v sobre o eixo que contém o vetor não-nulo u.

Se $z = \text{pr}_u(v)$, tem-se $v = z + w$, com $w \perp z$. Pelo Teorema de Pitágoras, $|v|^2 = |z|^2 + |w|^2$. Em particular vemos que $|z| \leq |v|$, isto é, o comprimento da projeção $\text{pr}_u(v)$ é menor do que ou igual ao comprimento de v.

Ora, a norma do vetor $\text{pr}_u(v)$ é igual a $|\langle u,v\rangle|/|u|$. Segue-se então que, para quaisquer $u, v \in E$, tem-se $|\langle u,v\rangle|/|u| \leq |v|$, ou seja

$$|\langle u,v\rangle| \leq |u| \cdot |v| \qquad \text{(\textit{desigualdade de Schwarz})}.$$

A rigor, o argumento acima prova a desigualdade de Schwarz apenas no caso em que $u \neq 0$. Mas ela é óbvia no caso em que $u = 0$. Logo vale em geral.

Um importante complemento da desigualdade de Schwarz é que vale a igualdade $|\langle u,v\rangle| = |u|\,|v|$ se, e somente se, um dos vetores u, v é múltiplo do outro. Isto resulta do raciocínio acima pois, no Teorema de Pitágoras $|v|^2 = |z|^2 + |w|^2$, dizer $|v| = |z|$ significa que $w = 0$, isto é, que v é múltiplo de u.

Resulta da desigualdade de Schwarz que num espaço vetorial com produto interno a norma satisfaz a *desigualdade triangular*:

$$|u + v| \leq |u| + |v|.$$

Como se trata de números não-negativos, para provar esta desigualdade basta mostrar que $|u+v|^2 \le (|u|+|v|)^2$. Ora,

$$\begin{aligned} |u+v|^2 &= \langle u+v, u+v\rangle \\ &= |u|^2 + |v|^2 + 2\langle u, v\rangle \\ &\le |u|^2 + |v|^2 + 2|u||v| \\ &= (|u|+|v|)^2, \end{aligned}$$

pois $\langle u, v\rangle \le |u|\,|v|$ pela desigualdade de Schwarz.

Vale a igualdade $|u+v| = |u|+|v|$ somente quando um dos vetores u, v é um múltiplo não-negativo do outro. Com efeito, pelo argumento acima, $|u+v| = |u|+|v|$ ocorre quando $\langle u, v\rangle = |u|\,|v|$, o que é óbvio quando $u = 0$ e implica $v = \alpha u$ quando $u \neq 0$. Neste caso, $|u|\,|v| = \langle u, v\rangle = \alpha\langle u, u\rangle = \alpha|u|^2$, logo $\alpha \ge 0$.

Além da desigualdade triangular, a norma goza ainda das seguintes propriedades, de imediata verificação:

$$|u| > 0 \quad \text{se} \quad u \neq 0 \quad \text{e} \quad |\alpha \cdot u| = |\alpha|\,|u|.$$

Em particular, $|-u| = |u|$.

Observação. Em todo este livro, $|\alpha|$ significa o valor absoluto do número α e $|u|$ representa também a norma do vetor u.

Num espaço vetorial E munido de produto interno, a *distância* entre os vetores u, v é, por definição, $d(u,v) = |u-v|$. Tem-se $d(u,u) = 0$, $d(u,v) > 0$ se $u \neq v$, $d(u,v) = d(v,u)$ e $d(u,w) \le d(u,v) + d(v,w)$.

Mostraremos agora que existem bases ortonormais em todo espaço vetorial de dimensão finita provido de um produto interno.

Mais precisamente, exporemos o *processo de ortonormalização de Gram-Schmidt*, um algoritmo que ensina a passar de uma base qualquer $\{v_1, \ldots, v_n\} \subset E$ para uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$, com a importante propriedade de que, para $m = 1, \ldots, n$, os vetores $u_1, \ldots, u_m$ pertencem ao subespaço F_m, gerado por $v_1, \ldots, v_m$.

Dada a base $\{v_1, \ldots, v_n\} \subset E$, obteremos primeiro uma base ortogonal $\{w_1, \ldots, w_n\} \subset E$ e depois poremos $u_1 = w_1/|w_1|, \ldots, u_n = w_n/|w_n|$ para chegar à base ortonormalizada $\{u_1, \ldots, u_n\} \subset E$.

Começamos o processo tomando $w_1 = v_1$ e prosseguimos por indução. Suponhamos já obtidos os vetores não-nulos $w_1, \ldots, w_m$,

dois a dois ortogonais, gerando o subespaço F_m, o mesmo que é gerado por $v_1, \ldots, v_m$. Definimos w_{m+1} pondo

$$w_{m+1} = v_{m+1} - \sum_{i=1}^{m} \frac{\langle w_i, v_{m+1} \rangle}{\langle w_i, w_i \rangle} w_i \,.$$

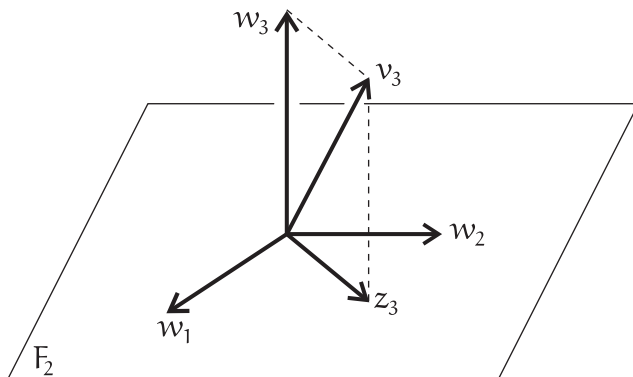



Figura 10.4. $w_3 = v_3 - z_3$, $z_3 = \frac{\langle w_1, v_3 \rangle}{\langle w_1, w_1 \rangle} w_1 + \frac{\langle w_2, v_3 \rangle}{\langle w_2, w_2 \rangle} w_2$.

Um cálculo simples mostra que w_{m+1} é ortogonal a $w_1, \ldots, w_m$. Além disso, $w_{m+1} \neq 0$ porque v_{m+1} não pertence ao subespaço F_m gerado por $w_1, \ldots, w_m$ (ou por $v_1, \ldots, v_m$). E, finalmente, w_{m+1} pertence ao subespaço gerado por $\{w_1, \ldots, w_m, v_{m+1}\}$, o qual é igual a F_{m+1}. Isto completa o processo.

Observamos que se os primeiros m vetores da base $\{v_1, \ldots, v_n\} \subset E$ já formarem uma base ortonormal do subespaço por eles gerado então o processo de Gram-Schmidt transforma essa base numa base ortonormal $\{u_1, \ldots, u_n\} \subset E$ na qual $u_1 = v_1, \ldots, u_m = v_m$.

Segue-se daí que, dado um subespaço vetorial $F \subset E$, toda base ortonormal de F estende-se a uma base ortonormal de E: basta estendê-la a uma base qualquer de E e depois ortonormalizar esta última por Gram-Schmidt.

O significado geométrico do processo de Gram-Schmidt é bastante simples e fica ilustrado na figura: se $\{w_1, \ldots, w_m\} \subset F$ é uma base ortogonal então, para todo $v \in E$, o vetor

$$z = \sum_{i=1}^{m} \frac{\langle w_i, v \rangle}{\langle w_i, w_i \rangle} \, w_i \in F,$$

soma das projeções ortogonais de v sobre os eixos dos w_i, tem a propriedade de que $w = v - z$ é perpendicular aos vetores $w_1, \ldots, w_m$. Daí resulta imediatamente que w é perpendicular a *todos* os vetores de F, pois esses vetores são combinações lineares dos w_i. O vetor z chama-se a *projeção ortogonal* de v sobre o subespaço F. Escreve-se $z = \text{pr}_F(v)$. (Fig. 10.5.) [Para a fórmula de $\text{pr}_F(v)$ quando a base $\{w_1, \ldots, w_m\} \subset F$ não é ortogonal, veja os Corolários 1 e 2 do Teorema 16.1 ou o Exercício 16.7.]

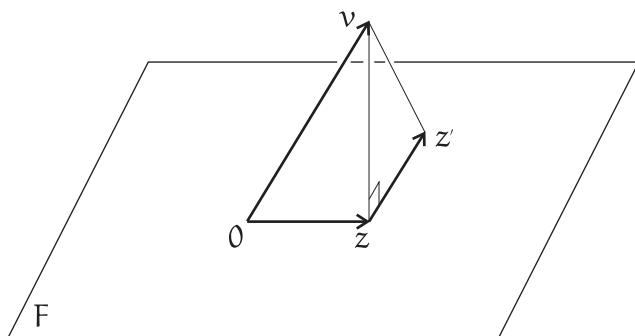



Figura 10.5.

Se z' é qualquer outro vetor em F, temos

$$v - z' = (v - z) + (z - z').$$

Como $z - z' \in F$, segue-se que $(v - z) \perp (z - z')$. Do Teorema de Pitágoras resulta então que $|v-z'|^2 = |v-z|^2 + |z-z'|^2$. Em particular, $|v - z'| \geq |v - z|$. Isto mostra que a distância de v à sua projeção $z = \text{pr}_F(v)$ sobre o subespaço F é menor do que ou igual à distância de v a qualquer outro vetor $z' \in F$. Noutras palavras, a projeção $z = \text{pr}_F(v)$ é o vetor mais próximo de v no subespaço F.

Se $\{u_1, \ldots, u_m\} \subset F$ é uma base ortonormal então a projeção ortogonal de um vetor $v \in E$ sobre o subespaço F se exprime, de forma mais simples, como

$$\text{pr}_F(v) = \sum_{i=1}^{m} \langle u_i, v \rangle \, u_i.$$

Isto está de acordo com a seguinte observação geral:

Seja $\{u_1, \ldots, u_n\} \subset E$ uma base ortonormal. Para todo vetor $v \in E$ tem-se

$$v = \sum_{i=1}^{n} \langle u_i, v\rangle \cdot u_i \,.$$

Com efeito, v se exprime como combinação linear

$$v = \alpha_1 u_1 + \cdots + \alpha_n u_n$$

em termos da base dada. Tomando o produto interno de ambos os membros desta igualdade por u_i, temos $\langle u_i, v\rangle = \alpha_i$ $(i = 1, \ldots, n)$ pois $\langle u_i, u_j\rangle = \delta_{ij}$ $(= 1$ se $i = j$ e $= 0$ se $i \neq j)$.

Assim, as coordenadas de um vetor relativamente a uma base ortonormal são os produtos internos desse vetor pelos elementos daquela base.

Se $u = \Sigma\alpha_i u_i$ e $v = \Sigma\beta_j u_j$ são as expressões dos vetores $u, v \in E$ em termos de uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$, as relações $\langle u_i, u_j\rangle = \delta_{ij}$ implicam imediatamente que

$$\langle u, v\rangle = \left\langle \sum_{i=1}^{n} \alpha_i u_i, \sum_{j=1}^{n} \beta_j u_j \right\rangle = \sum_{i,j=1}^{n} \alpha_i\beta_j \langle u_i, u_j\rangle = \sum_{i=1}^{n} \alpha_i\beta_i.$$

Portanto, quando se referem os vetores de E a uma base ortonormal fixada, o produto interno assume a forma $\langle u, v\rangle = \Sigma\alpha_i\beta_i$, análoga à do produto interno canônico de $\mathbb{R}^n$.

Tomando, mais geralmente, uma base arbitrária $\{v_1, \ldots, v_n\} \subset E$ e pondo $\langle v_i, v_j\rangle = g_{ij}$, o produto interno dos vetores

$$u = \sum_{i=1}^{n} \alpha_i v_i \quad \text{e} \quad v = \sum_{j=1}^{n} \beta_j v_j$$

se exprime como

$$\langle u, v\rangle = \sum_{i,j=1}^{n} g_{ij}\, \alpha_i\beta_j \,. \tag{*}$$

A matriz $\mathbf{g} = [g_{ij}] \in M(n \times n)$ é *simétrica*, isto é, $g_{ij} = g_{ji}$ pois $\langle v_i, v_j\rangle = \langle v_j, v_i\rangle$. Mais ainda: a matriz $\mathbf{g}$ é *positiva*. Isto significa que, além de $\mathbf{g}$ ser simétrica, para qualquer lista $(x_1, \ldots, x_n)$ de n números reais não todos nulos, tem-se

$$\sum_{i,j=1}^{n} g_{ij}\, x_i x_j > 0.$$

Reciprocamente, fixada uma base $\{v_1, \ldots, v_n\}$ num espaço vetorial E (que pode ser, por exemplo, a base canônica em $\mathbb{R}^n$) e dada uma matriz simétrica positiva $\mathbf{g} = [g_{ij}] \in M(n \times n)$, a igualdade (*) acima define um produto interno em E.

Exercícios

10.1. Seja E um espaço vetorial com produto interno. Para quaisquer vetores $u, v \in E$, prove que $|u|v + |v|u$ e $|u|v - |v|u$ são ortogonais.

10.2. Seja $\{u_1, \ldots, u_n\} \subset E$ uma base ortonormal. Prove que, para $v, w \in E$ arbitrários, tem-se $\langle v, w\rangle = \sum_{i=1}^{n} \langle v, u_i\rangle \langle w, u_i\rangle$.

10.3. Dado o vetor $u = (2, 3, 6)$, seja $P\colon \mathbb{R}^3 \to \mathbb{R}^3$ o operador linear definido por $Pv = \text{pr}_u(v)$. Descreva $I - 2P$ geometricamente, escreva a matriz da reflexão $H = I - 2P$ e determine o vetor que se obtém de $w = (1, 1, 1)$ por reflexão em torno do plano perpendicular a u.

10.4. Considere a base $\mathcal{V} = \{v_1, v_2, v_3\} \subset \mathbb{R}^3$, formada pelos vetores $v_1 = (1, 1, 1)$, $v_2 = (1, -1, 1)$ e $v_3 = (1, -1, -1)$. Determine a matriz de passagem $\mathbf{p}$, de $\mathcal{V}$ para a base ortonormal $\mathcal{U} = \{u_1, u_2, u_3\}$, obtida de $\mathcal{V}$ pelo método de Gram-Schmidt. Observe que os elementos da diagonal de $\mathbf{p}$ são números positivos e abaixo da diagonal todos são nulos. Generalize.

10.5. Seja $\mathcal{V} = \{v_1, \ldots, v_n\} \subset \mathbb{R}^n$ uma base, com $v_j = (\alpha_{1j}, \alpha_{2j}, \ldots, \ldots, \alpha_{nj})$, $j = 1, \ldots, n$. Seja $\mathcal{U}$ a base ortonormal de $\mathbb{R}^n$ obtida de $\mathcal{V}$ pelo processo de Gram-Schmidt. Prove que $\mathcal{U}$ é a base canônica de $\mathbb{R}^n$ se, e somente se, $\alpha_{ij} = 0$ para todo $i > j$ e $\alpha_{ii} > 0$ para todo $i = 1, \ldots, n$.

10.6. Sem fazer cálculo algum, diga quais são as bases obtidas de $\mathcal{V} = \{v_1, v_2, v_3\}$ pelo processo de Gram-Schmidt nos seguintes casos:

(a) $v_1 = (3, 0, 0)$, $v_2 = (-1, 3, 0)$, $v_3 = (2, -5, 1)$;

(b) $v_1 = (-1, 1, 0)$, $v_2 = (5, 0, 0)$, $v_3 = (2, -2, 3)$.

10.7. Dado o vetor unitário $u = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$, forme a matriz $\mathbf{a} = [\alpha_i \cdot \alpha_j] \in M(n \times n)$. Seja $H\colon \mathbb{R}^n \to \mathbb{R}^n$ o operador cuja matriz

na base canônica é $\mathbf{I}_n - 2\mathbf{a}$. Mostre que para todo $v \in \mathbb{R}^n$ tem-se $Hv = v - 2\langle v, u\rangle u$ e conclua que $|Hv| = |v|$.

10.8. Num espaço vetorial com produto interno, o ângulo entre dois vetores não-nulos u, v é, por definição, o ângulo $\theta = \measuredangle(u, v)$, tal que $0 \leq \theta \leq 180°$ e $\cos\theta = \langle u, v\rangle / |u|\,|v|$. Dito isto, e dados os vetores $u = (3, 4)$, $v = (1, -1)$ e $w = (-1, 1)$, ponha em ordem crescente os ângulos $\measuredangle(u, v)$, $\measuredangle(u, w)$ e $\measuredangle(v, w)$.

10.9. Sejam $u, v \in \mathbb{R}^2$ vetores L.I.. Prove que o vetor $|u|v + |v|u$ está contido na bissetriz do ângulo formado por u e v.

10.10. Seja $u = (a, b, c) \in \mathbb{R}^3$ um vetor unitário, com $abc \neq 0$. Determine t de modo que, pondo $v = (-bt, at, 0)$ e $w = (act, bct, -1/t)$, os vetores u, v, w sejam unitários e dois a dois ortogonais.

10.11. Para cada par de vetores $u = (x, y)$, $v = (x', y')$ em $\mathbb{R}^2$, ponha $[u, v] = 2xx' - xy' - x'y + 2yy'$. Prove que isto define um produto interno no espaço vetorial $\mathbb{R}^2$.

10.12. Dado o produto interno $\langle u, v\rangle$ no espaço vetorial E, prove que se tem $|u + v|^2 + |u - v|^2 = 2(|u|^2 + |v|^2)$ para quaisquer $u, v \in E$. Interprete esta igualdade geometricamente.

10.13. Seja X um conjunto de geradores do espaço vetorial E, onde está definido um produto interno. Se os vetores $u, v \in E$ são tais que $\langle u, w\rangle = \langle v, w\rangle$ para qualquer $w \in X$, prove que $u = v$.

10.14. Seja $\{v_1, \ldots, v_n\}$ uma base no espaço vetorial E, munido de produto interno. Dados n números reais arbitrários $\alpha_1, ..., \alpha_n$, prove que existe um, e somente um, vetor $w \in E$ tal que

$$\langle w, v_1\rangle = \alpha_1, \ldots, \langle w, v_n\rangle = \alpha_n.$$

10.15. Para toda base $\mathcal{V} = \{v_1, \ldots, v_n\}$ no espaço vetorial E, dotado de produto interno, prove que existe uma única base $\mathcal{W} = \{w_1, \ldots, w_n\} \subset E$ tal que $\langle w_i, v_j\rangle = \delta_{ij}$ $(i, j = 1, 2, \ldots, n)$. Se $\langle v_i, v_j\rangle = a_{ij}$ e $\langle w_i, w_j\rangle = b_{ij}$, prove que as matrizes $\mathbf{a} = [a_{ij}]$ e $\mathbf{b} = [b_{ij}]$ são inversas uma da outra.

10.16. Suponha que

$$[u, v] = \sum_{i,j=1}^{n} a_{ij} x_i y_j$$

defina, para $u = (x_1, \ldots, x_n)$ e $v = (y_1, \ldots, y_n)$, um produto interno em $\mathbb{R}^n$. Prove que $a_{11} > 0, \ldots, a_{nn} > 0$.

10.17. Calcule três produtos internos entre os vetores $u = (1, 0, -1)$, $v = (4, 1, 4)$, $w = (-3, 24, -3)$ e conclua que eles são linearmente independentes.

10.18. Em cada um dos casos abaixo, determine se o conjunto $\{u, v, w\} \subset \mathbb{R}^3$ é ortonormal, apenas ortogonal ou nenhum dos dois.

(a) $u = (1, 2, 1), v = (1, -1, 1), w = (-1, 1, 2)$.

(b) $u = (a, b, c), v = (-b, a, 0), w = (-ac, -bc, a^2 + b^2)$.

(c) $u = \frac{1}{7}(2, 6, 3), v = \frac{1}{7}(3, 2, -6), w = \frac{1}{7}(6, -3, 2)$.

10.19. Seja $\langle\ ,\ \rangle$ um produto interno no espaço vetorial F. Dado um isomorfismo $A\colon E \to F$, ponha $[u, v] = \langle Au, Av\rangle$ para quaisquer $u, v \in E$. Prove que $[\ ,\ ]$ é um produto interno em E.

10.20. Dados os vetores $u = (2, -1, 2)$, $v = (1, 2, 1)$ e $w = (-2, 3, 3)$, determine o vetor de $\mathbb{R}^3$ que é a projeção ortogonal de w sobre o plano gerado por u e v.

10.21. Qual é a base ortonormal de $\mathbb{R}^3$ obtida pelo processo de Gram-Schmidt a partir da base $\{u, v, w\}$, onde $u = (2, 6, 3)$, $v = (-5, 6, 24)$ e $w = (9, -1, -4)$?

10.22. Mesma pergunta do exercício anterior para $u = (3, 4, 12)$, $v = (7, -8, 15)$ e $w = (-15, 6, 44)$.

10.23. Para todo número natural n, prove que a norma do vetor $v = (n, n+1, n(n+1)) \in \mathbb{R}^3$ é um número natural.

10.24. Aplicando o processo de Gram-Schmidt a um conjunto de vetores $v_1, \ldots, v_m$ cuja independência linear não é conhecida, prove que se obtém o primeiro vetor $w_{r+1} = 0$ quando $v_1, \ldots, v_r$ são L.I. mas v_{r+1} é combinação linear de $v_1, \ldots, v_r$.

10.25. Fixado o vetor unitário $u = (a_1, \ldots, a_n) \in \mathbb{R}^n$, seja $P\colon \mathbb{R}^n \to \mathbb{R}^n$ o operador linear definido por $Pv = \mathrm{pr}_u(v) =$ projeção ortogonal de v sobre o eixo de u. Mostre que $P^2 = P$, determine o núcleo de P, as matrizes de P, de $I - P$ e da reflexão ortogonal $H = I - 2P$ em

torno do núcleo de P. (A matriz de H é conhecida como uma *matriz de Householder*.)

10.26. Seja a um vetor não-nulo no espaço vetorial E, de dimensão n, munido de produto interno. Para todo $b \in \mathbb{R}$, prove que o conjunto $V = \{v \in E; \langle v, a\rangle = b\}$ é uma variedade afim de dimensão $n-1$. Dado $v_o \in V$, mostre que $v \in V$ se, e somente se, $v - v_o$ é ortogonal a a.

10.27. Sejam $u = (x_1, x_2, x_3)$ e $v = (y_1, y_2, y_3)$ vetores em $\mathbb{R}^3$. O *produto vetorial* de u por v é definido como o vetor

$$u \times v = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1).$$

Prove que valem as seguintes propriedades:

(a) $u \times v = -v \times u$;

(b) $u \times (v + v') = u \times v + u \times v'$;

(c) $u \times (\alpha v) = \alpha(u \times v)$;

(d) $u \times v = 0$ se, e somente se, u e v são L.D.;

(e) $u \times v$ é ortogonal a u e a v;

(f) $e_1 \times e_2 = e_3$, $e_2 \times e_3 = e_1$, $e_3 \times e_1 = e_2$.

[Mais detalhes sobre o produto vetorial no livro "Coordenadas no Espaço", do autor, publicado pela Soc. Bras. de Mat.]

10.28. Seja $r = \{(1-t)u + tv; t \in \mathbb{R}\}$ a reta que liga u a v em E, com $u \neq v$. Dado $w \in E$, prove que, tomando $t = \langle w - u, v - u\rangle / |v - u|^2$ obtém-se o ponto $x = (1-t)u + tv$ de r mais próximo possível de w, ou seja, tem-se $|x - w| < |y - w|$ para qualquer outro ponto $y \in r$.

10.29. Seja $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ uma base no espaço vetorial E, munido de produto interno. Suponha que, para todo $v = x_1u_1 + \cdots + x_nu_n \in E$ se tenha $|v|^2 = x_1^2 + \cdots + x_n^2$. Prove que a base $\mathcal{U}$ é ortonormal.

10.30. Complete os detalhes do seguinte argumento que prova a existência de uma base ortonormal em qualquer espaço vetorial E, de dimensão n, com produto interno: "Seja $\mathcal{U} = \{u_1, \ldots, u_r\} \subset E$

um conjunto ortonormal com o maior número possível de elementos. Para todo vetor $v \in E$, o vetor

$$w = v - \sum_{i=1}^{r} \langle v, u_i \rangle \, u_i$$

é ortogonal a $u_1, \ldots, u_r$. Pela maximalidade de $\mathcal{U}$, tem-se $w = 0$, logo $\mathcal{U}$ gera E e é uma base ortonormal.”

10.31. Seja E um espaço vetorial com produto interno. Prove que, para quaisquer $u, v \in E$, tem-se $\big| |u| - |v| \big| \leq |u - v|$.

10.32. Prove que um operador $A \colon E \to E$, num espaço vetorial de dimensão finita com produto interno, tem posto 1 se, e somente se, existem vetores não-nulos $a, b \in E$ tais que $Av = \langle v, a \rangle \, b$ para todo $v \in E$. (Compare com o Exercício 8.27.)

10.33. Num espaço vetorial E com produto interno, o *cosseno* do ângulo entre dois vetores não-nulos u, v é definido como $\cos(u, v) = \langle u, v \rangle / |u| \, |v|$. Prove que se u e v são ortogonais e não-nulos então $\cos^2(u, u-v) + \cos^2(v, u-v) = 1$. (A soma dos quadrados dos cossenos dos ângulos agudos de um triângulo retângulo é igual a 1.)

10.34. Sejam E um espaço vetorial com produto interno, $C \subset E$ um conjunto convexo e $a \in E$ um ponto fora de C. Suponha que existam $x_o, x_1 \in C$ com a seguinte propriedade: para todo $x \in C$ tem-se $|a - x_o| \leq |a - x|$ e $|a - x_1| \leq |a - x|$. Prove que $x_o = x_1$.

11

A Adjunta

Mostraremos, nesta seção, como o produto interno nos permite associar a cada transformação linear $A\colon E \to F$ *uma nova transformação* $A^*\colon F \to E$*, chamada a adjunta de* A*. (Em espaços sem produto interno também existe uma noção de adjunta, mas aí se trata de uma transformação linear* $F^* \to E^*$*, do dual de* F *no dual de* E*. O produto interno nos dá condição de permanecer com* E *e* F*. Isto é particularmente interessante no caso de um operador linear* $A\colon E \to E$*.) A adjunta nos dá, por assim dizer, uma visão da transformação* A *sob um novo ângulo. Essa mudança de ponto de vista é reveladora, especialmente quando ocorre a existência de relações entre* A *e* A^**.*

Sejam $\dim E = n$ e $\dim F = m$. Vimos na Seção 8 que a escolha de bases em E e F determina um isomorfismo $\varphi\colon \mathcal{L}(E;F) \to M(m \times n)$, portanto o espaço vetorial $\mathcal{L}(E;F)$ das transformações lineares de E em F tem dimensão mn. Em particular, o espaço $E^* = \mathcal{L}(E;\mathbb{R})$ cujos elementos são os funcionais lineares $f\colon E \to \mathbb{R}$, chamado *espaço dual* de E, tem dimensão n. Isto implica que E^* é isomorfo a E. Na realidade, dada uma base $\mathcal{V} = \{v_1, \ldots, v_n\} \subset E$, existe uma base $\mathcal{V}^* = \{v_1^*, \ldots, v_n^*\} \subset E^*$, chamada *base dual* de $\mathcal{V}$, onde, por definição, para cada vetor $v = \Sigma\alpha_i v_i \in E$, tem-se $v_i^*(v) = \alpha_i$. (A verificação de que $\mathcal{V}^* \subset E^*$ é uma base pode ser feita diretamente ou então mediante a observação de que $\varphi(v_i^*) = e_i = i$-ésimo elemento da base canônica de $\mathbb{R}^n$, onde $\varphi\colon E^* \to M(1 \times n) = \mathbb{R}^n$ é o isomorfismo

acima mencionado.) Obtém-se um isomorfismo $A\colon E \to E^*$ impondo que $Av_i = v_i^*$ $(i = 1, \ldots, n)$.

Uma desvantagem dos isomorfismos entre E e E^* obtidos mediante o emprego de uma base é que eles não são intrínsecos: dado um vetor $v \in E$, o funcional $v^* \in E^*$ que a ele corresponde depende não apenas de v mas também da base de E que se tomou. Esta dificuldade, entretanto, desaparece quando E está munido de um produto interno, como veremos agora.

Seja E um espaço vetorial de dimensão finita, dotado de um produto interno. Definimos uma transformação linear $\xi\colon E \to E^*$ fazendo corresponder a cada vetor $v \in E$ o funcional linear $\xi \cdot v = v^*$, tal que $v^*(w) = \langle w, v\rangle$ para todo $w \in E$.

A verificação da linearidade de ξ é imediata: se $u, v \in E$, como

$$\begin{aligned}(u+v)^*(w) &= \langle w, u+v\rangle = \langle w, u\rangle + \langle w, v\rangle \\ &= u^*(w) + v^*(w) \\ &= [u^* + v^*](w)\end{aligned}$$

para todo $w \in E$, temos $(u+v)^* = u^* + v^*$. Analogamente, $(\alpha v)^* = \alpha \cdot v^*$.

Além disso, ξ é injetiva. Com efeito, dado $v \in E$, se $v^* = 0$ então, para todo $w \in E$ tem-se

$$\langle w, v\rangle = v^*(w) = 0(w) = 0.$$

Em particular, $\langle v, v\rangle = 0$, logo $v = 0$.

Finalmente, $\xi\colon E \to E^*$ é sobrejetiva pois é injetiva e os espaços E, E^* têm, como vimos, a mesma dimensão.

Assim, podemos enunciar o

Teorema 11.1. *Seja E um espaço vetorial de dimensão finita, com produto interno. A correspondência $\xi\colon E \to E^*$ que associa a cada $v \in E$ o funcional linear $\xi(v) = v^*$, tal que $v^*(w) = \langle w, v\rangle$ para todo $w \in E$, é um isomorfismo.*

O teorema acima será usado principalmente na medida em que assegura a existência de ξ^{-1}. Mais explicitamente: a todo funcional linear $f\colon E \to \mathbb{R}$ corresponde um único vetor $v = v_f \in E$ tal que $\langle w, v\rangle = f(w)$ para todo $w \in E$. Um tanto informalmente: para se conhecer um vetor $v \in E$ basta que se conheça o produto interno de

todos os vetores $w \in E$ por v (desde que esses produtos dependam linearmente de w).

O Teorema 11.1 é responsável pelo pouco (ou nenhum) uso que se faz de funcionais lineares em espaços, como $\mathbb{R}^n$, onde há um produto interno: funcionais são substituídos por vetores e a ação de um funcional sobre um vetor é substituída por um produto interno.

De posse do Teorema 11.1, definiremos a adjunta de uma transformação linear $A\colon E \to F$ onde E, F são espaços vetoriais de dimensão finita, ambos munidos de produto interno.

A *adjunta* de A deve ser uma transformação linear $A^*\colon F \to E$ tal que, para $v \in E$ e $w \in F$ quaisquer se tenha:

$$\langle Av, w\rangle = \langle v, A^*w\rangle\,. \qquad (*)$$

Assim, a imagem $A^*w \in E$ de um vetor arbitrário $w \in F$ é, por definição, aquele vetor de E tal que o produto interno de qualquer vetor $v \in E$ por ele é igual a $\langle Av, w\rangle$. Como, para cada $w \in F$, o número $f(v) = \langle Av, w\rangle$ depende linearmente de v (ou seja, f é um funcional linear), o Teorema 11.1 assegura que o vetor $A^*w \in E$ existe e é único de modo que valha (*) para quaisquer $v \in E$, $w \in F$. A correspondência $w \mapsto A^*w$ assim definida é uma transformação linear de F em E. Com efeito, dados $w, w' \in F$ tem-se, para todo $v \in E$:

$$\begin{aligned}\left\langle v, A^*(w+w')\right\rangle &= \left\langle Av, w+w'\right\rangle = \langle Av, w\rangle + \left\langle Av, w'\right\rangle \\ &= \langle v, A^*w\rangle + \left\langle v, A^*w'\right\rangle = \left\langle v, A^*w + A^*w'\right\rangle.\end{aligned}$$

Assim, $A^*(w+w')$ e $A^*w + A^*w'$ são vetores em E cujos produtos internos por qualquer vetor $v \in E$ são iguais. Portanto $A^*(w+w') = A^*w + A^*w'$. De modo análogo se verifica que $A^*(\alpha w) = \alpha \cdot A^*w$. Assim, $A^* \in \mathcal{L}(F;E)$.

A *transposta* de uma matriz $\mathbf{a} = [a_{ij}] \in M(m \times n)$ é a matriz $\mathbf{a}^T = [a_{ji}] \in M(n \times m)$ que tem como linhas as colunas de $\mathbf{a}$ e como colunas as linhas de $\mathbf{a}$, na mesma ordem.

Teorema 11.2. *Sejam $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ e $\mathcal{V} = \{v_1, \ldots, v_m\} \subset F$ bases ortonormais. Se $\mathbf{a} = [a_{ij}] \in M(m \times n)$ é a matriz da transformação linear $A\colon E \to F$ nas bases $\mathcal{U}$, $\mathcal{V}$ então a matriz da adjunta $A^*\colon F \to E$ nas bases $\mathcal{V}$, $\mathcal{U}$ é a transposta $\mathbf{a}^T = [a_{ji}] \in M(n \times m)$ de $\mathbf{a}$.*

Demonstração: Por definição de matriz de uma transformação linear, temos

$$Au_j = \sum_{i=1}^{m} a_{ij} v_i \quad (j = 1, \ldots, n)$$

e

$$A^* v_i = \sum_{r=1}^{n} b_{ri} u_r,$$

onde $\mathbf{b} = [b_{ri}] \in M(n \times m)$ é a matriz de A^* nas bases $\mathcal{V}$, $\mathcal{U}$, a ser determinada. Como ambas as bases são ortonormais, temos, para cada $i = 1, \ldots, m$ e cada $j = 1, \ldots, n$:

$$b_{ji} = \langle u_j, A^* v_i \rangle = \langle Au_j, v_i \rangle = a_{ij}$$

portanto, $\mathbf{b} = \mathbf{a}^T$, transposta de $\mathbf{a}$. □

Corolário. *Uma transformação linear A e sua adjunta A^* têm o mesmo posto.* (Vide Teorema 8.2.)

É apresentada a seguir uma lista de propriedades operacionais da adjunta de uma transformação linear, as quais se traduzem em propriedades da transposta de uma matriz, via Teorema 11.2. A validez dessas propriedades decorre da observação de que duas transformações lineares $A, B \colon E \to F$ são iguais quando se tem $\langle Au, v \rangle = \langle Bu, v \rangle$ para quaisquer $u \in E$, $v \in F$.

$I^* = I$	$(\mathbf{I}_n)^T = \mathbf{I}_n$
$(A + B)^* = A^* + B^*$	$(\mathbf{a} + \mathbf{b})^T = \mathbf{a}^T + \mathbf{b}^T$
$(\alpha A)^* = \alpha A^*$	$(\alpha \mathbf{a})^T = \alpha \mathbf{a}^T$
$(BA)^* = A^* B^*$	$(\mathbf{ba})^T = \mathbf{a}^T \mathbf{b}^T$
$A^{**} = A$	$(\mathbf{a}^T)^T = \mathbf{a}$

Se $A \colon E \to F$ é uma transformação linear injetiva então existe $B \colon F \to E$ tal que $BA = I_E$ (vide Teorema 6.5). Tomando a adjunta de ambos os membros desta igualdade, temos $A^* B^* = I_E$. Assim $A^* \colon F \to E$ possui uma inversa à direita B^*, logo é sobrejetiva. (Teorema 6.1.) Do mesmo modo se vê que A sobrejetiva implica A^* injetiva. Portanto a adjunta de um isomorfismo $A \colon E \to F$ é um isomorfismo $A^* \colon F \to E$. Além disso, de $A^{-1} A = I_E$ resulta $A^* (A^{-1})^* = I_E$ logo $(A^*)^{-1} = (A^{-1})^*$.

Analogamente, uma matriz quadrada $\mathbf{a}$ é invertível se, e somente se, sua transposta $\mathbf{a}^T$ é invertível e, no caso afirmativo, $(\mathbf{a}^T)^{-1} = (\mathbf{a}^{-1})^T$.

As noções de retas e planos perpendiculares da Geometria Elementar se estendem em Álgebra Linear ao conceito de complemento ortogonal, o qual ajuda a entender as relações entre uma transformação linear e sua adjunta.

Seja E um espaço vetorial com produto interno. O *complemento ortogonal* de um conjunto não-vazio $X \subset E$ é o conjunto $X^\perp$ formado pelos vetores $v \in E$ que são ortogonais a todos os vetores $x \in X$. Portanto

$$v \in X^\perp \Leftrightarrow \langle v, x \rangle = 0 \quad \text{para todo} \quad x \in X.$$

- Dado $X \subset E$, temos $\langle 0, x \rangle = 0$ para todo $x \in X$, logo $0 \in X^\perp$.
- Se $v \in X^\perp$ e $\alpha \in \mathbb{R}$ então $\langle \alpha v, x \rangle = \alpha \langle v, x \rangle = 0$ para todo $x \in X$, portanto $\alpha v \in X^\perp$.
- Se $u \in X^\perp$ e $v \in X^\perp$ então, para todo $x \in X$, tem-se $\langle u + v, x \rangle = \langle u, x \rangle + \langle v, x \rangle = 0$, logo $u + v \in X^\perp$.

Segue-se das três observações acima que o complemento ortogonal de qualquer conjunto não-vazio $X \subset E$ é um subespaço vetorial de E.

Evidentemente, $X \subset Y \Rightarrow Y^\perp \subset X^\perp$ e $v \in X \cap X^\perp \Rightarrow v = 0$.

Além disso, se v é ortogonal aos vetores $x_1, \ldots, x_m$ então v é ortogonal a qualquer combinação linear $\Sigma \alpha_i x_i$, pois

$$\left\langle v, \sum \alpha_i x_i \right\rangle = \sum \alpha_i \langle v, x_i \rangle = 0.$$

Daí resulta que o complemento ortogonal $X^\perp$ do conjunto X coincide com o complemento ortogonal $S(X)^\perp$ do subespaço vetorial $S(X)$, gerado por X.

Exemplo 11.1. Tem-se $\{0\}^\perp = E$ e $E^\perp = \{0\}$. Se $F \subset \mathbb{R}^n$ é o subespaço vetorial gerado pelo vetor não nulo $v = (a_1, \ldots, a_n)$ (reta que passa pela origem), o complemento ortogonal $F^\perp$ é o hiperplano definido pela equação $a_1 x_1 + \cdots + a_n x_n = 0$.

Teorema 11.3. *Seja E um espaço vetorial de dimensão finita, munido de produto interno. Para todo subespaço vetorial $F \subset E$ tem-se a decomposição em soma direta $E = F \oplus F^\perp$.*

Demonstração: Seja $\{u_1, \ldots, u_n\} \subset E$ uma base ortonormal cujos primeiros m elementos $u_1, \ldots, u_m$ formam uma base (ortonormal) de F. (Começa-se com uma base qualquer de F, estende-se-a a uma base de E e depois aplica-se Gram-Schmidt.) Para todo vetor $v \in E$ tem-se $v = \alpha_1 u_1 + \cdots + \alpha_n u_n = z + w$, onde $z = \alpha_1 u_1 + \cdots + \alpha_m u_m \in F$ e $w = \alpha_{m+1} u_{m+1} + \cdots + \alpha_n u_n \in F^\perp$. Portanto $E = F + F^\perp$. Como $F \cap F^\perp = \{0\}$, segue-se que $E = F \oplus F^\perp$. □

Corolário 1. dim $F +$ dim $F^\perp =$ dim E.

Corolário 2. *Para todo subespaço vetorial* $F \subset E$*, tem-se* $(F^\perp)^\perp = F$.

Com efeito, seja qual for o conjunto não-vazio $X \subset E$, vale a inclusão $X \subset (X^\perp)^\perp$. Em particular, o subespaço F está contido em $(F^\perp)^\perp$. Do Corolário 1 resulta que

$$\dim(F^\perp)^\perp = \dim E - \dim F^\perp = \dim E - (\dim E - \dim F) = \dim F.$$

Logo $F = (F^\perp)^\perp$. □

Vimos na Seção 10 que a projeção ortogonal de um vetor $v \in E$ sobre um subespaço $F \subset E$ é, por definição, o vetor

$$z = \mathrm{pr}_F(v) = \sum_{i=1}^{m} \frac{\langle w_i, v\rangle}{\langle w_i, w_i\rangle} w_i\,,$$

onde $\{w_1, \ldots, w_m\} \subset F$ é uma base ortogonal. Vimos ainda que, pondo $w = v - z$, temos $v = z + w$, com $z \in F$ e w perpendicular a todos os w_i $(i = 1, \ldots, m)$, logo $w \in F^\perp$. Ficou no ar a questão de saber até que ponto o vetor $z = \mathrm{pr}_F(v)$ depende da escolha da base ortogonal $\{w_1, \ldots, w_m\} \subset F$. A resposta é dada pelo Teorema 11.3. Como $E = F \oplus F^\perp$, é *única* a maneira de escrever um vetor $v \in E$ como soma $v = z + w$ de um vetor $z \in F$ com um vetor $w \in F^\perp$. Isto mostra que $z = \mathrm{pr}_F(v)$ não depende da escolha dos w_i. (Veja também o Exercício 11.5.)

De agora em diante, indicaremos com a notação $P_F \colon E \to E$, ou simplesmente $P \colon E \to E$, quando não houver perigo de confusão, a projeção associada à decomposição $E = F \oplus F^\perp$, a qual chamaremos a *projeção ortogonal* sobre F.

Teorema 11.4. *Dada a transformação linear* $A: E \to F$*, entre espaços vetoriais de dimensão finita munidos de produto interno, tem-se*

$$\begin{aligned} \mathcal{N}(A^*) &= \mathcal{I}m(A)^\perp, \\ \mathcal{I}m(A^*) &= \mathcal{N}(A)^\perp, \\ \mathcal{N}(A) &= \mathcal{I}m(A^*)^\perp \quad \text{e} \\ \mathcal{I}m(A) &= \mathcal{N}(A^*)^\perp. \end{aligned}$$

Demonstração: Basta provar a primeira dessas igualdades; as demais se seguem dela usando $A^{**} = A$ e $F^{\perp\perp} = F$. Ora,

$$\begin{aligned} v \in \mathcal{N}(A^*) \Leftrightarrow A^*v = 0 \Leftrightarrow \langle u, A^*v \rangle = 0 \text{ q.q.s. } u \in E \Leftrightarrow \\ \Leftrightarrow \langle Au, v \rangle = 0 \text{ q.q.s. } u \in E \Leftrightarrow v \in \mathcal{I}m(A)^\perp. \end{aligned}$$

□

Corolário 1. *A fim de que o sistema de* m *equações lineares com* n *incógnitas*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad (i = 1, \ldots, m)$$

possua solução é necessário e suficiente que o vetor $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$ *seja perpendicular a toda solução* $y = (y_1, \ldots, y_m)$ *do sistema homogêneo transposto*

$$\sum_{j=1}^{m} a_{ji}y_j = 0 \quad (i = 1, \ldots, n).$$

Com efeito, pela última das igualdades do Teorema 11.4, o sistema $Ax = b$ tem solução se, e somente se, b é ortogonal ao núcleo de A^*, isto é, a todas as soluções $y \in \mathbb{R}^m$ do sistema homogêneo $A^*y = 0$. □

O ponto do Corolário 1 é que ele permite concluir a existência de soluções sem que seja necessário exibir uma delas.

Corolário 2. *O posto de* A^* *é igual ao posto de* A.

Com efeito, se $\dim E = n$ então $\dim \mathcal{N}(A) + \dim \mathcal{I}m(A) = n$, logo

$$\begin{aligned}\dim \mathcal{I}m(A^*) &= n - \dim \mathcal{N}(A)\\ &= n - [n - \dim \mathcal{I}m(A)]\\ &= \dim \mathcal{I}m(A).\end{aligned}$$

□

Esta prova do Corolário 2 é uma alternativa para o corolário do Teorema 11.2 sem o uso de matrizes.

Na Fig. 11.1, os pares de retas perpendiculares representam pares de subespaços, cada um dos quais é o complemento ortogonal do outro.

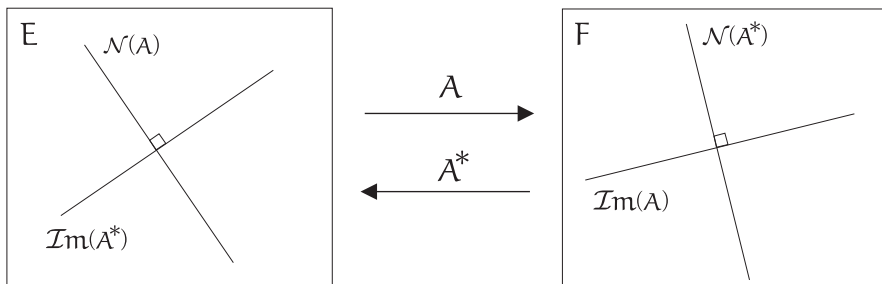



Figura 11.1.

Exercícios

11.1. Seja $A\colon E \to F$ uma transformação linear entre espaços vetoriais de dimensão finita, munidos de produto interno. Prove:

(a) Se A é sobrejetiva então $AA^*\colon F \to F$ é invertível e $A^*(AA^*)^{-1}\colon F \to E$ é uma inversa à direita de A.

(b) Se A é injetiva então $A^*A\colon E \to E$ é invertível e $(A^*A)^{-1}A^*$ é uma inversa à esquerda de A.

11.2. Use o exercício anterior a fim de achar uma inversa à direita para a transformação linear $A\colon \mathbb{R}^3 \to \mathbb{R}^2$, dada por $A(x, y, z) = (x + 2y + 3z, 2x - y - z)$ e uma inversa à esquerda para a transformação linear $B\colon \mathbb{R}^2 \to \mathbb{R}^4$, onde $A(x, y) = (x + 2y, 2x - y, x + 3y, 4x + y)$.

11.3. Dada a matriz $\mathbf{a} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, calcule $\mathbf{aa}^T$ e, a partir daí, encontre uma matriz $\mathbf{b} \in M(3 \times 2)$ tal que $\mathbf{ab} = \mathbf{I}_2$.

11.4. Seja $P\colon E \to E$ uma projeção num espaço vetorial de dimensão finita, munido de produto interno. Prove que P^* também é uma projeção. Dê um exemplo em que $P^* \neq P$.

11.5. Seja $\mathcal{U} = \{u_1, \ldots, u_r\}$ uma base ortonormal do subespaço F, contido no espaço vetorial E, dotado de produto interno. Usando $\mathcal{U}$, defina a aplicação $P\colon E \to E$, pondo, para todo $v \in E$, $Pv = \mathrm{pr}_u(v) = \sum_{i=1}^{r} \langle v, u_i \rangle\, u_i$. Prove que P é um operador linear com $\mathcal{I}m(P) = F$, $\mathcal{N}(P) = F^\perp$ e $P^2 = P$. Obtenha assim outra demonstração de que $E = F \oplus F^\perp$. (Vide Teorema 7.2.)

11.6. Considere, no espaço vetorial $M(n \times n)$, o produto interno definido por

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i,j} a_{ij} b_{ij},$$

se $\mathbf{a} = [a_{ij}]$ e $\mathbf{b} = [b_{ij}]$. (Vide Exercício 11.17.) Mostre que o subespaço A das matrizes anti-simétricas é o complemento ortogonal do subespaço S das matrizes simétricas em $M(n \times n)$.

11.7. No espaço vetorial E das funções contínuas $f\colon [-1, 1] \to \mathbb{R}$, sejam $F, G \subset E$ os subespaços vetoriais formados pelas funções pares e pelas funções ímpares, respectivamente. Relativamente ao produto interno $\langle f, g \rangle = \int_{-1}^{1} f(x) g(x)\, dx$, em E, mostre que G é o complemento ortogonal de F.

11.8. Se os operadores lineares $A, B\colon E \to E$ comutam (isto é, $AB = BA$), prove que A^* e B^* também comutam.

11.9. Sejam $\{u_1, \ldots, u_n\} \subset E$ e $\{v_1, \ldots, v_m\} \subset F$ conjuntos de geradores nesses espaços vetoriais com produto interno. Sejam ainda $A\colon E \to F$ e $B\colon F \to E$ transformações lineares tais que $\langle Au_j, v_i \rangle = \langle u_j, Bv_i \rangle$ para $i = 1, \ldots, m$ e $j = 1, \ldots, n$. Prove que $B = A^*$.

11.10. Dada a matriz $\mathbf{a} \in M(m \times n)$, prove que ou o sistema $\mathbf{ax} = \mathbf{b}$ tem solução qualquer que seja $\mathbf{b} \in M(m \times 1)$ ou o sistema homogêneo transposto $\mathbf{a}^T\mathbf{y} = \mathbf{0}$ admite uma solução não-trivial.

11.11. No espaço $M(n \times n)$, munido do produto interno $\langle \mathbf{a}, \mathbf{b} \rangle = \text{tr}(\mathbf{a}^T\mathbf{b})$, (veja Exercício 11.17) considere uma matriz fixa $\mathbf{a}$ e defina o operador linear $T_\mathbf{a}: M(n \times n) \to M(n \times n)$ pondo $T_\mathbf{a}\mathbf{x} = \mathbf{ax}$. Mostre que a adjunta de $T_\mathbf{a}$ é $T_\mathbf{b}$, onde $\mathbf{b} = \mathbf{a}^T$. Prove um resultado análogo para o operador $S_\mathbf{a}: M(n \times n) \to M(n \times n)$, onde $S_\mathbf{a}\mathbf{x} = \mathbf{xa}$. (Obs. $\text{tr}(\mathbf{ab}) = \text{tr}(\mathbf{ba})$.)

11.12. Seja $S: \mathbb{R}^3 \to \mathbb{R}^3$ a reflexão em torno do plano $z = 0$, paralelamente à reta $x = y = z$. Determine a adjunta S^*. Mesma questão para a projeção $P: \mathbb{R}^3 \to \mathbb{R}^3$, sobre o mesmo plano, paralelamente à mesma reta.

11.13. Sejam $A, B: E \to E$ operadores lineares num espaço vetorial de dimensão finita, munido de produto interno. Prove que se $B^*A = 0$ então, para todo $v \in E$, os vetores Av e Bv são perpendiculares. Em particular, se $A^*A = 0$ então $A = 0$.

11.14. Para todo conjunto não-vazio X num espaço vetorial munido de produto interno, prove que $X^{\perp\perp} = S(X)$ (subespaço vetorial gerado por X).

11.15. Sejam F_1, F_2 subespaços do espaço E, munido de produto interno. Prove que $(F_1 + F_2)^\perp = F_1^\perp \cap F_2^\perp$, $(F_1 \cap F_2)^\perp = F_1^\perp + F_2^\perp$.

11.16. Dê mais uma demonstração de que A e A^* têm o mesmo posto, nas seguintes linhas: a afirmação é verdadeira quando $A: E \to F$ é injetiva ou sobrejetiva pois nestes casos A^* é sobrejetiva ou injetiva, respectivamente. Para a conclusão no caso geral, use o fato de que toda transformação linear se escreve como um produto BA, onde B é injetiva e A é sobretiva. (Exercício 6.29.)

11.17. Sejam E, F espaços vetoriais de dimensão finita, munidos de produto interno. Dadas as transformações lineares $A, B: E \to F$, ponha $\langle A, B \rangle = \text{tr}(A^*B)$ e prove que isto define um produto interno em $\mathcal{L}(E; F)$. (Vide Exercício 8.37.) Se $\mathbf{a} = [a_{ij}]$ e $\mathbf{b} = [b_{ij}]$ são as matrizes de A e B em relação a bases ortonormais de E e F respectivamente, prove que $\langle A, B \rangle = \sum_{i,j} a_{ij}b_{ij}$.

11.18. Prove que uma projeção $P: E \to E$, num espaço com produto interno, é ortogonal (isto é, $\mathcal{N}(P) = \mathcal{I}m(P)^{\perp}$) se, e somente se, para todo $v \in E$ tem-se $\langle Pv, v - Pv \rangle = 0$.

11.19. Use o exercício anterior para provar que se uma projeção $P: E \to E$ cumpre $|Pv| \leq |v|$ para todo $v \in E$ então P é ortogonal. [Sugestão: suponha que, para algum $v \in E$, Pv não fosse ortogonal a $v - Pv$. Tome w = pé da perpendicular baixada de 0 sobre a reta que contém v e Pv. Então $|w| < |Pv|$. Mas todos os pontos desta reta têm a mesma imagem por P. Logo $|Pv| = |Pw|$ e daí $|w| < |Pw|$, uma contradição.]

11.20. Ache uma base para o complemento ortogonal do subespaço (plano) de $\mathbb{R}^3$ gerado pelos vetores $u = (3, -1, 2)$ e $v = (-1, 2, 3)$.

11.21. Dado o operador $A: \mathbb{R}^3 \to \mathbb{R}^3$, definido por $A(x, y, z) = (x + y + z, 3x - 2y - z, -2x + 3y + 2z)$, obtenha bases para os seguintes subespaços de $\mathbb{R}^3$: $\mathcal{I}m(A)$, $\mathcal{N}(A)$, $\mathcal{I}m(A^*)$ e $\mathcal{N}(A^*)$.

11.22. Considere a base $\{u, v, w\} \subset \mathbb{R}^3$ onde $u = (1, 1, 1)$, $v = (1, 2, 3)$, $w = (1, -2, 1)$. Determine as matrizes (na base canônica) dos funcionais lineares $u^*, v^*, w^*: \mathbb{R}^3 \to \mathbb{R}$ que formam a base dual de $\{u, v, w\}$.

11.23. Com a notação do exercício anterior, a base $\{u, v, w\} \subset \mathbb{R}^3$ determina um isomorfismo $\psi: (\mathbb{R}^3)^* \to \mathbb{R}^3$, que a cada funcional $f \in (\mathbb{R}^3)^*$ faz corresponder sua matriz (do tipo 1×3) na base dada. Prove que $\psi(f) = [a, b, c] \Leftrightarrow f = au^* + bv^* + cw^*$.

11.24. Demonstre que $(AB)^* = B^*A^*$ e $A^{**} = A$.

11.25. Estabeleça uma conexão entre os Exercícios 4.20 e 10.15 por intermédio do Teorema 11.1.

11.26. Seja $A: E \to F$ uma transformação linear entre espaços de dimensão finita, com produto interno. Se $\dim E < \dim F$, prove que o operador $AA^*: F \to F$ não é invertível mas se $\mathcal{N}(A) = \{0\}$ então $A^*A: E \to E$ é invertível. Dê um exemplo desta situação com $E = \mathbb{R}^2$ e $F = \mathbb{R}^3$. Que se pode afirmar quando $\dim E > \dim F$?

11.27. Num espaço vetorial E munido de produto interno, sejam $\mathcal{V} = \{v_1, \ldots, v_n\}$ e $\mathcal{W} = \{w_1, \ldots, w_n\}$ bases tais que $\langle v_i, w_j \rangle = \delta_{ij}$. (Cfr. Exercício 10.15.) Prove que a matriz do operador linear A^* na base $\mathcal{W}$ é a transposta da matriz do operador A na base $\mathcal{V}$.

11.28. Seja $\mathbf{a}$ uma matriz quadrada. Se o traço (soma dos elementos da diagonal) de $\mathbf{a}^T.\mathbf{a}$ é zero, prove que $\mathbf{a} = \mathbf{0}$.

11.29. Uma matriz quadrada $\mathbf{a}$ chama-se *diagonalizável* quando é semelhante a uma matriz $\mathbf{d} = [d_{ij}]$ do tipo diagonal ($d_{ij} = 0$ se $i \neq j$), ou seja, quando existe $\mathbf{p}$ invertível tal que $\mathbf{p}^{-1}\mathbf{ap} = \mathbf{d}$. Prove que se $\mathbf{a}$ é diagonalizável então $\mathbf{a}^T$ também o é. Se a matriz do operador $A\colon E \to E$ relativamente a uma base de E é diagonalizável, prove que a matriz de A em relação a qualquer outra base é diagonalizável.

11.30. Prove que a adjunta de $A\colon E \to E$, onde $Av = \langle v, a\rangle\, b$, é $A^*v = \langle v, b\rangle\, a$.

11.31. Seja $f^*\colon \mathbb{R} \to E$ a adjunta do funcional linear $f\colon E \to \mathbb{R}$. Prove que $v = f^*(1)$ é o vetor de E que corresponde a f pelo isomorfismo do Teorema 11.1. Prove ainda que $f(f^*(1)) = |v|^2$ e $f^*(f(w)) = \langle w, v\rangle v \quad \forall w \in E$.

11.32. Para toda transformação linear $A\colon E \to F$, entre espaços de dimensão finita munidos de produto interno, prove que a restrição de A à imagem de A^* define um isomorfismo $A\colon \mathcal{I}m(A^*) \to \mathcal{I}m(A)$. Analogamente, A^* transforma o subespaço $\mathcal{I}m(A)$ isomorficamente sobre $\mathcal{I}m(A^*)$. São estes isomorfismos um o inverso do outro?

11.33. Seja $\{u_1, \ldots, u_n\} \subset E$ uma base ortonormal. Para todo operador linear $A\colon E \to E$, prove que

$$\sum_{i=1}^{n} |Au_i|^2 = \sum_{i=1}^{n} |A^*u_i|^2.$$

11.34. Seja $J\colon F \to E$ a *inclusão* do subespaço $F \subset E$, isto é, $Jv = v$ para todo $v \in F$. Prove que $J^*J\colon F \to F$ é o operador identidade de F e que o operador $JJ^*\colon E \to E$ é a projeção de núcleo $F^\perp$ e imagem F. (Noutras palavras, a adjunta $J^*\colon E \to F$ é esta mesma projeção, porém considerada com contra-domínio F.)

12

Subespaços Invariantes

Quanto menor é a dimensão do espaço E*, mais fácil é estudar os operadores lineares* $A\colon E \to E$*. (Isto é especialmente verdadeiro quando* $\dim E = 1$ *ou* $\dim E = 2$*.) Por isso, quando se tem um operador* $A\colon E \to E$*, é natural que se tente, de alguma maneira, "decompô-lo" em operadores definidos em subespaços de dimensões menores. O passo inicial nessa busca é a noção de subespaço invariante por um operador, que estudaremos nesta seção. E o caso de maior êxito é o dos operadores auto-adjuntos; como veremos na seção seguinte, todo operador daquele tipo se decompõe em operadores uni-dimensionais. Outros exemplos de sucesso serão vistos nas Seções 14 e 15. As bases serão lançadas agora.*

Provaremos nesta seção (Teorema 12.1) que, dado um operador linear $A\colon E \to E$ num espaço vetorial de dimensão finita, ou bem existe um vetor não-nulo $u \in E$ tal que $Au = \lambda u$ ou então existem $u, v \in E$ linearmente independentes tais que Au e Av são ambos combinações lineares de u e v: $Au = \alpha u + \beta v$, $Av = \gamma u + \delta v$. Este fato será fundamental para o estudo de certos tipos particularmente importantes de operadores, como os auto-adjuntos e os ortogonais, que abordaremos nas seções seguintes.

Para demonstrar o Teorema 12.1, faremos uso do chamado Teorema Fundamental da Álgebra, do qual resulta que todo polinômio mônico real se decompõe como produto de polinômios mônicos irre-

dutíveis do primeiro e do segundo graus. (Lembramos que se chama *mônico* um polinômio no qual o coeficiente do termo de mais alto grau é igual a 1 e que um polinômio irredutível do segundo grau não admite raiz real.)

O teorema a ser demonstrado significa que existe em E um subespaço vetorial de dimensão 1 ou 2, invariante por A de acordo com a seguinte definição.

Diz-se que um subespaço vetorial $F \subset E$ é *invariante* pelo operador linear $A\colon E \to E$ quando $A(F) \subset F$, isto é, quando a imagem Av de qualquer vetor $v \in F$ é ainda um vetor em F.

Se F é um subespaço invariante do operador $A\colon E \to E$, a restrição de A aos vetores de F define um operador que, salvo quando houver perigo de confusão, indicaremos com a mesma notação $A\colon F \to F$. Assim, a existência de um subespaço invariante permite o estudo de um operador mais simples, por estar definido num domínio menor.

Exemplo 12.1. Os subespaços $\{0\}$ e E são invariantes por qualquer operador $A\colon E \to E$. O núcleo $\mathcal{N}(A)$ e a imagem $\mathcal{I}m(A)$ são também exemplos óbvios de subespaços invariantes. Um subespaço F de dimensão 1 (reta passando pela origem) é invariante por A se, e somente se, existe um número λ tal que $Av = \lambda v$ para todo $v \in F$. [Com efeito, fixando um vetor $u \neq 0$ em F, todos os demais elementos de F são da forma αu, $\alpha \in \mathbb{R}$. Como $Au \in F$, tem-se $Au = \lambda u$. Para qualquer outro $v \in F$, vale $v = \alpha u$ logo $Av = \alpha Au = \alpha\lambda u = \lambda\alpha u$, logo $Av = \lambda v$, com o mesmo λ.] Se $u, v \in E$ são linearmente independentes, o subespaço F gerado por u e v (plano contendo a origem) é invariante por A se, e somente se $Au \in F$ e $Av \in F$, isto é, $Au = \alpha u + \beta v$ e $Av = \gamma u + \delta v$.

Um vetor $v \neq 0$ em E chama-se um *autovetor* do operador $A\colon E \to E$ quando existe $\lambda \in \mathbb{R}$ tal que

$$Av = \lambda v.$$

O número $\lambda \in \mathbb{R}$, por sua vez, chama-se um *autovalor* do operador A quando existe um vetor não-nulo $v \in E$ tal que $Av = \lambda v$. Diz-se então que o autovalor λ corresponde, ou pertence, ao autovetor v e, vice-versa, que o autovetor v também corresponde, ou pertence, ao autovalor λ. Então, para todo $w = \alpha v$, tem-se $Aw = \lambda w$.

Achar um autovetor (ou, o que é equivalente, um autovalor) do operador A é, portanto, o mesmo que achar um subespaço de dimensão 1 invariante por A.

Analogamente, diz-se que o número real λ é um *autovalor da matriz* $\mathbf{a} \in M(n \times n)$ quando λ é um autovalor do operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$, cuja matriz na base canônica é $\mathbf{a}$. Isto significa que existe um vetor $x \neq 0$ em $\mathbb{R}^n$ tal que $Ax = \lambda x$ ou, o que é, o mesmo, uma matriz não-nula $\mathbf{x} \in M(n \times 1)$ tal que $\mathbf{ax} = \lambda \mathbf{x}$.

Exemplo 12.2. Uma rotação $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ em torno da origem, de ângulo diferente de $0°$ e $180°$, não admite outros subespaços invariantes além de $\{0\}$ e $\mathbb{R}^2$. Por outro lado, para todo $\alpha \in \mathbb{R}$, a rotação $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ de ângulo α em torno do eixo z, definida por

$$A(x, y, z) = (x \cos \alpha - y \operatorname{sen} \alpha, x \operatorname{sen} \alpha + y \cos \alpha, z),$$

tem o eixo z e o plano $z = 0$ como subespaços invariantes. Para todo $z \neq 0$, o vetor $v = (0, 0, z)$ é um autovetor de A, cujo autovalor correspondente é 1, pois $Av = v$. Já no caso de uma reflexão $S\colon E \to E$ em torno do subespaço F_1, paralelamente a F_2 (vide Teorema 7.3), todo vetor não-nulo em F_1 é um autovetor de S, com autovalor 1, enquanto que os vetores não-nulos em F_2 são autovetores correspondentes ao autovalor -1. Finalmente, se o operador A tem núcleo não-trivial então todo vetor não-nulo $v \in \mathcal{N}(A)$ é um autovetor pois $Av = 0 \cdot v$.

Exemplo 12.3. O operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, definido por $A(x, y) = (x + \alpha y, y)$, chama-se *cisalhamento*. Se $\alpha \neq 0$, os únicos subespaços invariantes por A são $\{0\}$, $\mathbb{R}^2$ e o eixo das abcissas. Com efeito, qualquer outro subespaço de $\mathbb{R}^2$ é uma reta F, formada pelos múltiplos $tv = (ta, tb)$ de um vetor $v = (a, b)$, com $b \neq 0$. Se $t \neq 0$ tem-se $tv \in F$ mas $A(tv) = (ta + \alpha tb, tb) = tv + (\alpha tb, 0) \notin F$ logo F não é invariante por A.

Dados o polinômio $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ e o operador $A\colon E \to E$, a notação $p(A)$ indica o operador

$$p(A) = a_0 I + a_1 A + \cdots + a_n A^n.$$

Lema. *Para todo operador linear* $A\colon E \to E$*, num espaço vetorial de dimensão finita, existem um polinômio mônico irredutível* p*, de grau* 1 *ou* 2*, e um vetor não-nulo* $v \in E$*, tais que* $p(A) \cdot v = 0$.

Demonstração: Seja $n = \dim E$. Como $\dim \mathcal{L}(E) = n^2$, os $n^2 + 1$ operadores $I, A, \ldots, A^{n^2}$ são linearmente dependentes. Portanto existem números reais $\alpha_o, \ldots, \alpha_{n^2}$, não todos nulos, tais que $\alpha_o I + \alpha_1 A + \cdots + \alpha_{n^2} A^{n^2} = 0$. Seja α_m o coeficiente não-nulo de maior índice nesta expressão. Dividindo-a por α_m, obtemos um polinômio mônico $p(x) = \beta_o + \beta_1 x + \cdots + \beta_{m-1} x^{m-1} + x^m$ tal que $p(A) = 0$. Sabemos que existe uma fatoração $p(x) = p_1(x) \cdot p_2(x) \cdots p_k(x)$, onde cada p_i é um polinômio mônico irredutível de grau 1 ou 2. Temos $p_1(A) \cdot p_2(A) \cdots p_k(A) = 0$. Logo, pelo menos um dos operadores $p_i(A)$ não é invertível. Assim, existe um vetor não-nulo $v \in E$ tal que $p_i(A) \cdot v = 0$. □

Teorema 12.1. *Todo operador linear num espaço vetorial de dimensão finita possui um subespaço invariante de dimensão 1 ou 2.*

Demonstração: Dado $A \colon E \to E$, sejam p o polinômio e $v \in E$ o vetor não-nulo dados pelo lema, com $p(A) \cdot v = 0$. Se $p(x) = x - \lambda$ então $Av - \lambda v = 0$, donde $Av = \lambda v$, logo a reta que passa pela origem e contém v é um subespaço invariante por A, de dimensão 1. Se p tem grau 2, $p(x) = x^2 + ax + b$, então $A^2 v + aAv + bv = p(A) \cdot v = 0$, logo $A(Av) = -aAv - bv$. Isto mostra que o subespaço gerado por v e Av é invariante por A. Além disso, v e Av são L.I. pois se tivéssemos $Av = \lambda v$ então

$$0 = A^2 v + aAv + bv = \lambda^2 v + a\lambda v + bv = (\lambda^2 + a\lambda + b)v,$$

donde $\lambda^2 + a\lambda + b = 0$, uma contradição pois $p(x) = x^2 + ax + b$ não tem raiz real. Logo o subespaço invariante gerado por v e Av tem dimensão 2. □

Um operador linear num espaço vetorial de dimensão n admite no máximo n autovalores distintos. Isto é conseqüência do

Teorema 12.2. *A autovalores diferentes do mesmo operador correspondem autovetores linearmente independentes.*

Demonstração: Dado o operador linear $A \colon E \to E$, sejam $v_1, \ldots, v_m$ vetores não-nulos em E tais que $Av_1 = \lambda_1 v_1, \ldots, Av_m = \lambda_m v_m$, onde os números reais $\lambda_1, \ldots, \lambda_m$ são dois a dois diferentes. Provaremos, por indução, que esses vetores são L.I.. A afirmação é óbvia quando $m = 1$. Supondo-a verdadeira para $m - 1$ vetores, inferiremos daí

sua validez para m. Dada a combinação linear nula

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0, \tag{*}$$

aplicamos o operador A a ambos os membros desta igualdade, levando em conta que $Av_i = \lambda_i v_i$. Resulta então que

$$\lambda_1 \alpha_1 v_1 + \cdots + \lambda_m \alpha_m v_m = 0. \tag{**}$$

Multiplicando a igualdade (*) por λ_m e subtraindo de (**) vem:

$$(\lambda_1 - \lambda_m)\alpha_1 v_1 + \cdots + (\lambda_{m-1} - \lambda_m)\alpha_{m-1} v_{m-1} = 0.$$

Pela hipótese de indução, os vetores $v_1, \ldots, v_{m-1}$ são L.I.. Logo

$$(\lambda_1 - \lambda_m)\alpha_1 = \cdots = (\lambda_{m-1} - \lambda_m)\alpha_{m-1} = 0.$$

Como os autovalores são todos diferentes, as $m - 1$ diferenças nos parênteses acima são $\neq 0$, logo $\alpha_1 = \cdots = \alpha_{m-1} = 0$. Isto reduz a igualdade (*) a $\alpha_m v_m = 0$. Como $v_m \neq 0$, segue-se que $\alpha_m = 0$. Assim, a igualdade (*) só pode ocorrer quando todos os coeficientes α_i são nulos, o que prova o teorema. □

Corolário. *Seja* $\dim E = n$. *Se um operador linear* $A\colon E \to E$ *possui* n *autovalores diferentes então existe uma base* $\{v_1, \ldots, v_n\} \subset E$ *em relação à qual a matriz de* A *é diagonal (isto é, tem a forma* $[a_{ij}]$ *com* $a_{ij} = 0$ *se* $i \neq j$*).*

Com efeito, se $Av_1 = \lambda_1 v_1, \ldots, Av_n = \lambda_n v_n$ com os v_i não-nulos e os λ_i dois a dois distintos então $\{v_1, \ldots, v_n\}$ é, em virtude do Teorema 12.2, uma base de E. A matriz de A nesta base é

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

na qual os termos que não aparecem são iguais a zero. □

A igualdade $Av = \lambda v$ equivale a $(A - \lambda I)v = 0$, logo v é um autovetor do operador $A\colon E \to E$ se, e somente se, é um elemento não-nulo do núcleo $\mathcal{N}(A - \lambda I)$. Noutras palavras, a fim de que λ

seja um autovalor de A é necessário e suficiente que o operador $A - \lambda I \colon E \to E$ não possua inverso.

Exemplo 12.4. Um caso particular importante ocorre quando $\dim E = 2$. Vimos no Exemplo 2.6 que se $\{u, v\} \subset E$ é uma base então os vetores $\alpha u + \beta v$ e $\gamma u + \delta v$ são linearmente dependentes se, e somente se, $\alpha\delta - \beta\gamma = 0$. Dados o operador $A \colon E \to E$ e a base $\{u, v\} \subset E$, sejam $Au = au + cv$ e $Av = bu + dv$. Noutras palavras, a matriz do operador A na base $\{u, v\}$ é

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Então $(A - \lambda I)u = (a - \lambda)u + cv$ e $(A - \lambda I)v = bu + (d - \lambda)v$. A fim de que $A - \lambda I$ não seja invertível é necessário e suficiente que os vetores $(A - \lambda I)u$ e $(A - \lambda I)v$ sejam L.D., ou seja, que $(a - \lambda)(d - \lambda) - bc = 0$, ou ainda, que λ seja raiz do polinômio

$$p(\lambda) = \lambda^2 - (a + d)\lambda + ad - bc,$$

chamado o *polinômio característico* do operador A.

Portanto, o número real λ é um autovalor do operador $A \colon E \to E$, onde $\dim E = 2$, se, e somente se, é uma raiz do polinômio característico do operador A, o qual, por definição, é $p(\lambda) = \lambda^2 - (a + d)\lambda + ad - bc$. Os coeficientes de $p(\lambda)$ são tirados da matriz de A em relação a uma base qualquer de E.

Observação. A matriz do operador A muda quando se passa de uma base para outra. Mas o polinômio $p(\lambda)$ (isto é, as expressões $a + d$ e $ad - bc$, que são seus coeficientes) permanece (isto é, permanecem) sem alteração. Isto será provado na Seção 20. No presente caso ($\dim E = 2$), é claro que $a + d$ = traço de A, logo independe da base escolhida. Quanto ao coeficiente $ad - bc$ (o determinante da matriz $\mathbf{a} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$), vide o Exercício 12.7.

Exemplo 12.5. No caso da rotação $R \colon \mathbb{R}^2 \to \mathbb{R}^2$, $R(x, y) = (x \cos\theta - y \operatorname{sen}\theta, x \operatorname{sen}\theta + y \cos\theta)$, temos $a = \cos\theta$, $b = -\operatorname{sen}\theta$, $c = \operatorname{sen}\theta$, $d = \cos\theta$, logo o polinômio característico de R é

$$p(\lambda) = \lambda^2 - (2\cos\theta)\lambda + 1.$$

Se $\theta \neq 0$ e $\theta \neq 180°$, o trinômio $p(\lambda)$ não possui raiz real pois neste caso seu discriminante $\Delta = 4(\cos^2\theta - 1)$ é negativo. Conseqüentemente R só possui autovalores (reais) se $\theta = 0$ ou $\theta = 180°$.

Exemplo 12.6. Definamos o operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ pondo $A(x, y) = (4x + 3y, x + 2y)$. Seu polinômio característico é $p(\lambda) = \lambda^2 - 6\lambda + 5$, cujas raízes são $\lambda_1 = 1$ e $\lambda_2 = 5$. Estes números são autovalores de A. Existem, portanto, vetores não-nulos v_1 e v_2 em $\mathbb{R}^2$, tais que $Av_1 = v_1$ e $Av_2 = 5v_2$. Pelo Teorema 12.2, v_1 e v_2 formam uma base de $\mathbb{R}^2$, em relação à qual a matriz do operador A tem a forma diagonal:

$$\mathbf{a} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

A fim de determinar os vetores $v_1 = (x, y)$ e $v_2 = (r, s)$ exprimimos as igualdades $Av_1 = v_1$ e $Av_2 = 5v_2$ em termos de coordenadas, obtendo os sistemas lineares

$$\begin{aligned} 4x + 3y &= x \\ x + 2y &= y \end{aligned}$$

e

$$\begin{aligned} 4r + 3s &= 5r \\ r + 2s &= 5s. \end{aligned}$$

Ambos os sistemas acima são indeterminados, e tinham que ser assim pois se v é autovetor de A, todo múltiplo αv também é. Tomando uma solução não-nula de cada um desses sistemas obtemos $v_1 = (1, -1)$, $v_2 = (3, 1)$ tais que $\{v_1, v_2\} \subset \mathbb{R}^2$ é uma base formada por autovetores de A.

Exercícios

12.1. Suponha que o operador linear $A\colon E \to E$, num espaço vetorial de dimensão n, admita um subespaço invariante F, de dimensão r. Prove que existe uma base de E, relativamente à qual a matriz de A tem a forma

$$\begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} \end{bmatrix},$$

com $\mathbf{a} \in M(r \times r)$, $\mathbf{b} \in M(r \times (n-r))$, $\mathbf{0} \in M((n-r) \times r)$ e $\mathbf{c} \in M((n-r) \times (n-r))$.

12.2. Enuncie e prove um resultado análogo ao do exercício anterior, supondo que $E = F \oplus G$, onde F e G são invariantes pelo operador A.

12.3. Dê exemplo de um operador linear $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ que admite um subespaço invariante $F \subset \mathbb{R}^3$ com a seguinte propriedade: nenhum subespaço $G \subset \mathbb{R}^3$ tal que $\mathbb{R}^3 = F \oplus G$ é invariante por A.

12.4. Dado o vetor não-nulo $a \in \mathbb{R}^3$, determine os subespaços de $\mathbb{R}^3$ invariantes pelo operador $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, definido por $Av = a \times v$ (produto vetorial: veja Exercício 10.27).

12.5. Sejam $A, B\colon E \to E$ operadores lineares. Se $AB = BA$, prove que $\mathcal{N}(B)$ e $\mathcal{I}m(B)$ são subespaços invariantes por A.

12.6. Dado o operador linear $A\colon E \to E$ e o polinômio $p(x)$, prove que os subespaços vetoriais $\mathcal{N}(p(A))$ e $\mathcal{I}m(p(A))$ são invariantes por A.

12.7. Um operador $A\colon E \to E$ chama-se *normal* quando $AA^* = A^*A$. Prove que se A é normal então, para todo $v \in E$, tem-se $|Av| = |A^*v|$ e conclua daí que todo auto-vetor de A é também auto-vetor de A^*, com o mesmo auto-valor. (Sugestão: se A é normal, $A-\lambda I$ também é.)

12.8. Se o operador A é normal, prove que $\mathcal{N}(A)^\perp = \mathcal{I}m(A)$.

12.9. Seja $P\colon \mathbb{R}^3 \to \mathbb{R}^3$ uma projeção de posto 2. Prove que os únicos subespaços de $\mathbb{R}^3$ invariantes por P estão contidos na imagem ou contêm o núcleo de P. E se o posto de P for igual a 1?

12.10. Mostre que os subespaços vetoriais de $C^\infty(\mathbb{R}, \mathbb{R})$ gerados por cada um dos conjuntos abaixo são invariantes pelo operador de derivação $D\colon C^\infty(\mathbb{R}, \mathbb{R}) \to C^\infty(\mathbb{R}, \mathbb{R})$.

(a) $\{\cos x, \operatorname{sen} x\}$;

(b) $\{e^x, xe^x, x^2e^x\}$.

12.11. Se $F \subset \mathbb{R}^3$ é um subespaço de dimensão 2, invariante pelo operador linear $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, e A não possui auto-vetores em F, prove que nenhum subespaço de dimensão 2, além de F, é invariante por A.

12.12. Dado o operador linear $A\colon E \to E$ num espaço vetorial de dimensão 3, prove que existe uma base de E relativamente à qual a

matriz de A tem uma das formas abaixo:

$$\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ e & f & g \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} a & b & e \\ c & d & f \\ 0 & 0 & g \end{bmatrix}.$$

Em qualquer caso, prove que existe um vetor não-nulo $v \in E$ tal que $Av = gv$. [Observação: veremos na Seção 20 que, mais geralmente, todo operador linear num espaço vetorial de dimensão ímpar possui pelo menos um auto-vetor.]

12.13. Se $F_1, F_2 \subset E$ são subespaços invariantes pelo operador linear $A \colon E \to E$, prove que $F_1 \cap F_2$ e $F_1 + F_2$ também são invariantes por A.

12.14. Seja $A \colon E \to E$ um operador invertível, com $\dim E$ finita. Se o subespaço $F \subset E$ é invariante por A, prove que a restrição de A a F é ainda um operador invertível $F \to F$ e conclua que F é também invariante por A^{-1}.

12.15. Quais são os autovetores e autovalores do operador de derivação $D \colon \mathcal{P} \to \mathcal{P}$?

12.16. Num espaço vetorial E de dimensão n, prove que todo operador linear possui um subespaço invariante de dimensão $n-1$ ou $n-2$. (Sugestão: aplique o Teorema 12.1 à adjunta do operador dado, tendo antes introduzido um produto interno em E.)

12.17. Se v e w são respectivamente autovetores de A e A^*, correspondentes a autovalores $\lambda \neq \mu$, prove que $\langle v, w \rangle = 0$.

12.18. Prove que um operador A é invertível se, e somente se, não tem autovalor igual a 0. No caso afirmativo, prove que os autovetores de A e de A^{-1} coincidem. E os autovalores?

12.19. O *determinante* da matriz $\mathbf{a} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ é, por definição, o número $\det \mathbf{a} = ad - bc$. Mediante um cálculo direto, mostre que se $\mathbf{m} = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ então $\det(\mathbf{am}) = \det \mathbf{a} \cdot \det \mathbf{m}$. Prove ainda que $\det \mathbf{a} \neq 0$ se, e somente se, $\mathbf{a}$ é invertível. Conclua que, para toda matriz invertível $\mathbf{m}$, tem-se $\det \mathbf{a} = \det(\mathbf{m}^{-1}\mathbf{am})$, logo todas as matrizes do operador $A \colon E \to E$, com $\dim E = 2$, têm o mesmo determinante, o qual é chamado o *determinante do operador* A.

12.20. Dado o operador linear $A: E \to E$, suponha que $E = F_1 \oplus F_2$, onde $Av_1 = \lambda_1 v_1$ se $v_1 \in F_1$ e $Av_2 = \lambda_2 v_2$ se $v_2 \in F_2$, com $\lambda_1 \neq \lambda_2$. Prove que λ_1 e λ_2 são os únicos autovalores de A e que os autovetores de A estão em F_1 ou em F_2.

12.21. Seja $A: \mathbb{R}^n \to \mathbb{R}^n$ o operador linear cuja matriz na base canônica tem todos os elementos iguais a 1. Prove que o posto de A é igual a 1 e que $\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{I}m(A)$. Conclua que os autovalores de A são 0 e n e que seus autovetores pertencem a $\mathcal{N}(A)$ ou a $\mathcal{I}m(A)$. Exiba uma base de $\mathbb{R}^n$ na qual a matriz de A tem $n^2 - 1$ zeros.

12.22. Se todo vetor não-nulo de E for um autovetor do operador linear $A: E \to E$, prove que $A = \lambda I$.

12.23. Para todo autovalor λ do operador linear $A: E \to E$, seja $E_\lambda = \{v \in E; Av = \lambda v\}$. Prove que E_λ é um subespaço vetorial de E, invariante por A. (E_λ chama-se *auto-subespaço* correspondente ao autovalor λ.)

12.24. Prove que um subespaço que contém o núcleo de uma projeção é invariante por essa projeção.

12.25. Se $AB = BA$, prove que a imagem por B de um subespaço invariante por A é ainda invariante por A.

12.26. Seja $A: E \to E$ um operador linear tal que A^2 possui algum autovalor ≥ 0. Prove que A possui autovetor. Dê um exemplo em que A^2 possui autovetor mas A não possui.

12.27. Se os autovetores do operador linear $A: E \to E$ geram o espaço E e, além disso, os subespaços invariantes por A são também invariantes por B, prove que $AB = BA$.

12.28. Seja $\dim E = n$. Se o operador linear $A: E \to E$ possui n autovalores distintos, prove que existem no espaço E exatamente 2^n subespaços invariantes por A.

12.29. Seja $A: E \to E$ um operador no espaço vetorial E, de dimensão finita, onde $E = F_1 \oplus \cdots \oplus F_k$ e cada F_i é invariante por A. Tome uma base $\mathcal{V} \subset E$ que seja uma união de bases dos F_i. Determine a forma da matriz de A na base $\mathcal{V}$.

12.30. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador definido por $A(x,y) = (y,0)$. Quais são os autovalores de A? E os autovetores? Se $\mathbf{a} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, existe alguma matriz invertível $\mathbf{p} \in M(2\times 2)$ tal que $\mathbf{p}^{-1}\mathbf{ap}$ seja uma matriz diagonal?

12.31. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador definido por $A(x,y) = (2x - y, x + 4y)$. Mostre que A possui um autovalor único igual a 3 e que o auto-subespaço E_3 (v. Exercício 12.23) tem dimensão 1. Conclua que se

$$\mathbf{a} = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix}$$

então não existe uma matriz invertível $\mathbf{b} \in M(2 \times 2)$ tal que $\mathbf{b}^{-1}\mathbf{ab}$ seja diagonal.

12.32. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador definido por $A(x,y) = (3x + y, 2x+2y)$. Mostre que A possui os autovalores 4 e 1. Ache uma base $\{u,v\} \subset \mathbb{R}^2$ tal que $Au = 4u$ e $Av = v$. Dada a matriz

$$\mathbf{a} = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix},$$

ache uma matriz invertível $\mathbf{p} \in M(2 \times 2)$ tal que $\mathbf{p}^{-1}\mathbf{ap} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$.

12.33. Prove que todo operador linear de posto 1 em $\mathbb{R}^n$ possui um autovetor cujo autovalor correspondente é o traço do operador dado.

12.34. Seja p um polinômio $\neq 0$ cujas raízes são todas reais. Se $p(A) = 0$, prove que pelo menos uma raiz de p é autovalor do operador A.

12.35. Se o espaço vetorial E possui uma base formada por autovetores do operador $A\colon E \to E$, prove que existe também uma base de E formada por autovetores de $A^*\colon E \to E$. (Veja Exercício 11.29.)

12.36. Sejam $A\colon E \to E$ um operador linear num espaço de dimensão finita, munido de produto interno e $F \subset E$ um subespaço invariante por A. Defina o operador $B\colon F \to F$ pondo $Bv = Av$. Mostre que, para todo $v \in F$, se tem $A^*v = B^*v + Cv$, com $Cv \in F^{\perp}$.

13

Operadores Auto-Adjuntos

O Teorema Espectral para operadores auto-adjuntos, a ser provado nesta seção, é um dos resultados mais relevantes da Álgebra Linear. Serão também demonstradas algumas de suas conseqüências, entre as quais se destaca o Teorema dos Valores Singulares.

Um operador linear $A\colon E \to E$, num espaço vetorial munido de produto interno, chama-se *auto-adjunto* quando $A = A^*$, ou seja, quando $\langle Au, v\rangle = \langle u, Av\rangle$ para quaisquer $u, v \in E$.

Se $A, B\colon E \to E$ são operadores auto-adjuntos e $\alpha \in \mathbb{R}$ então $(A + B)^* = A^* + B^* = A + B$ e $(\alpha A)^* = \alpha A^* = \alpha A$, logo $A + B$ e αA são auto-adjuntos.

O produto AB dos operadores auto-adjuntos A, B é auto-adjunto se, e somente se, A e B comutam, isto é, $AB = BA$. Com efeito, sendo A e B auto-adjuntos, temos

$$(AB)^* = B^*A^* = BA.$$

Logo, AB é auto-adjunto se, e somente se, $BA = AB$.

Exemplo 13.1. Sejam $A, B\colon \mathbb{R}^2 \to \mathbb{R}^2$ os operadores lineares definidos por $A(x, y) = (x, 2y)$ e $B(x, y) = (y, x)$. Para todo $v = (x, y)$ tem-se:

$$\langle e_1, A^*v\rangle = \langle Ae_1, v\rangle = \langle e_1, v\rangle = x$$
$$\langle e_2, A^*v\rangle = \langle Ae_2, v\rangle = \langle 2e_2, v\rangle = 2y,$$

portanto $A^*v = (x, 2y) = Av$ e $A^* = A$. Analogamente se mostra que $B^* = B$. Entretanto, como $AB(x, y) = (y, 2x)$, vê-se que, para $v = (x, y)$, $\langle e_1, (AB)^*v\rangle = \langle ABe_1, v\rangle = 2y$, enquanto $\langle e_1, ABv\rangle = y$, logo $(AB)^* \neq AB$, ou seja, o produto AB dos operadores auto-adjuntos A, B não é auto-adjunto. Isto se dá porque $AB \neq BA$. Com efeito, $AB(x, y) = (y, 2x)$ e $BA(x, y) = (2y, x)$.

Exemplo 13.2. A projeção ortogonal $P\colon E \to E$ sobre um subespaço $F \subset E$ é um operador auto-adjunto. Com efeito, dados $v = z + w$, $v' = z' + w'$, com $z, z' \in F$ e $w, w' \in F^\perp$, temos

$$\left\langle Pv, v'\right\rangle = \left\langle z, v'\right\rangle = \left\langle z, z'\right\rangle = \left\langle v, z'\right\rangle = \left\langle v, Pv'\right\rangle.$$

Reciprocamente, se a projeção $P\colon E \to E$ sobre o subespaço F_1, paralelamente a F_2, onde $E = F_1 \oplus F_2$, for um operador auto-adjunto então para quaisquer $v_1 \in F_1$, $v_2 \in F_2$ vale

$$\langle v_1, v_2\rangle = \langle Pv_1, v_2\rangle = \langle v_1, Pv_2\rangle = \langle v_1, 0\rangle = 0,$$

logo $F_2 = F_1^\perp$. Assim, a projeção $P\colon E \to E$ é um operador auto-adjunto se, e somente se, é uma projeção ortogonal.

Uma matriz quadrada $\mathbf{a} = [a_{ij}]$ diz-se *simétrica* quando é igual à sua transposta $\mathbf{a}^T$, isto é, quando $a_{ij} = a_{ji}$ para todo i e todo j.

No teorema 13.1 é dado um operador linear $A\colon E \to E$, num espaço vetorial de dimensão finita, dotado de produto interno.

Teorema 13.1. $A\colon E \to E$ *é auto-adjunto se, e somente se, sua matriz* $\mathbf{a} = [a_{ij}]$ *relativamente a uma (e portanto a qualquer) base ortonormal* $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ *é uma matriz simétrica.*

Demonstração: $\langle u_i, Au_j\rangle$ = [i-ésima coordenada do vetor Au_j na base $\mathcal{U}$] = [i-ésimo elemento da j-ésima coluna de $\mathbf{a}$] = a_{ij}. Portanto a matriz $\mathbf{a}$ é simétrica se, e somente se, $\langle u_i, Au_j\rangle = \langle Au_i, u_j\rangle$ para quaisquer $i, j = 1, \ldots, n$. Devido à linearidade de A e à bilinearidade do produto interno, isto equivale a dizer que $\langle u, Av\rangle = \langle Au, v\rangle$ para quaisquer $u, v \in E$, ou seja, que A é auto-adjunto. □

Exemplo 13.3. As matrizes dos operadores A e B do Exemplo 13.1 na base canônica de $\mathbb{R}^2$ são respectivamente

$$\mathbf{a} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

ambas simétricas. Quanto ao Exemplo 13.2, se tomarmos em E uma base ortonormal cujos primeiros m elementos formem uma base de F e os últimos uma base de $F^\perp$, a matriz da projeção P nesta base terá os m primeiros termos da diagonal iguais a 1 e todos os demais elementos iguais a zero. Seu formato será

$$\begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix},$$

onde os termos fora da diagonal, não indicados acima, são todos zeros. Essas matrizes são simétricas, refletindo o fato de que representam operadores auto-adjuntos em bases ortonormais. Já o operador A no Exemplo 12.6 não é auto-adjunto pois na base canônica de $\mathbb{R}^2$ sua matriz é $\begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$.

Teorema 13.2 *Seja* $A\colon E \to E$ *um operador auto-adjunto. Se o subespaço* $F \subset E$ *é invariante por* A*, seu complemento ortogonal* $F^\perp$ *também é.*

O Teorema 13.2 decorre imediatamente do

Teorema 13.3 *Se o subespaço* $F \subset E$ *é invariante pelo operador linear* $A\colon E \to E$ *então seu complemento ortogonal* $F^\perp$ *é invariante pelo operador adjunto* $A^*\colon E \to E$.

Demonstração: $[u \in F, v \in F^\perp] \Rightarrow Au \in F \Rightarrow \langle u, A^*v\rangle = \langle Au, v\rangle = 0 \Rightarrow A^*v \in F^\perp$, logo $F^\perp$ é invariante por A^*. $\square$

Exemplo 13.4. No cisalhamento $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, onde $A(x,y) = (x + \alpha y, y)$, com $\alpha \neq 0$, o eixo x, das abcissas, é invariante mas seu complemento ortogonal, o eixo y, das ordenadas, não é pois $Ae_2 = (\alpha, 1)$ não é vertical.

No caso de um operador auto-adjunto, o Teorema 12.2 assume a forma mais precisa seguinte.

Teorema 13.4. *Se $\lambda_1, \ldots, \lambda_m$ são autovalores dois a dois diferentes do operador auto-adjunto $A\colon E \to E$, os autovetores correspondentes $v_1, \ldots, v_m$ são dois a dois ortogonais.*

Demonstração: Para $i \neq j$ quaisquer:

$$\begin{aligned}(\lambda_i - \lambda_j)\langle v_i, v_j\rangle &= \langle \lambda_i v_i, v_j\rangle - \langle v_i, \lambda_j v_j\rangle = \langle Av_i, v_j\rangle - \langle v_i, Av_j\rangle \\ &= \langle Av_i, v_j\rangle - \langle Av_i, v_j\rangle = 0 \text{ pois } A \text{ é auto-adjunto.}\end{aligned}$$

Como $\lambda_i - \lambda_j \neq 0$, de $(\lambda_i - \lambda_j)\langle v_i, v_j\rangle = 0$ resulta $\langle v_i, v_j\rangle = 0$. □

Observação. Se $Av = \lambda v$ então, para todo múltiplo $w = \alpha v$, tem-se ainda $Aw = \lambda w$. Logo, na situação do Teorema 13.4, os vetores $v_1, \ldots, v_m$ podem ser tomados unitários, caso haja conveniência.

Um problema importante sobre operadores num espaço vetorial de dimensão finita é o de encontrar uma base em relação à qual a matriz desse operador seja a mais simples possível. Mostraremos nesta seção que, se $A\colon E \to E$ é um operador auto-adjunto num espaço vetorial de dimensão finita com produto interno, existe uma base ortonormal em E, relativamente à qual a matriz de A é uma *matriz diagonal* $\mathbf{a} = [a_{ij}]$, isto é $a_{ij} = 0$ se $i \neq j$. Este é o conteúdo do Teorema Espectral.

Existe um tipo de operador auto-adjunto para o qual o Teorema Espectral é imediato: se $P\colon E \to E$ é a projeção ortogonal sobre o subespaço F, tomando uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$ cujos primeiros vetores $u_1, \ldots, u_m$ formem uma base de F (portanto os $n - m$ últimos formam uma base de $F^\perp$), a matriz de P nesta base tem a forma diagonal vista no Exemplo 13.3.

Quando se diz que a matriz do operador $A\colon E \to E$ na base $\{u_1, \ldots, u_n\} \subset E$ é uma matriz diagonal, isto significa que, para todo $j = 1, \ldots, n$, tem-se $Au_j = \lambda_j u_j$, ou seja, que os vetores da base dada são todos eles autovetores de A.

No caso da projeção ortogonal sobre o subespaço F, tem-se $Pu_j = u_j$ para $j = 1, \ldots, m$ e $Pu_j = 0$ se $j = m + 1, \ldots, n$. Assim, a base ortonormal acima fixada é de fato formada por autovetores de P. Os autovalores são 1 e 0.

Começamos com o caso particular do Teorema Espectral em que espaço tem dimensão 2.

Teorema 13.5. *Seja* $A\colon E \to E$ *um operador auto-adjunto num espaço vetorial de dimensão 2, munido de produto interno. Existe uma base ortonormal* $\{u_1, u_2\} \subset E$ *formada por autovetores de* A.

Demonstração: Seja $\{v, w\} \subset E$ uma base ortonormal arbitrária. Em virtude do Teorema 13.1, temos $Av = av + bw$, $Aw = bv + cw$. Como vimos antes no Exemplo 12.4, os autovalores de A são as raízes reais do polinômio característico $p(\lambda) = \lambda^2 - (a+c)\lambda + ac - b^2$. O discriminante deste trinômio é $\Delta = (a+c)^2 - 4(ac - b^2) = (a-c)^2 + 4b^2 \geq 0$. Se $\Delta = 0$ então $b = 0$, $a = c$ e $A = aI$, logo todo vetor não-nulo em E é um autovetor. Se $\Delta > 0$ então o trinômio $p(\lambda)$ possui 2 raízes reais distintas λ_1, λ_2. Isto, como sabemos, quer dizer que os operadores $A - \lambda_1 I$ e $A - \lambda_2 I$ são ambos não invertíveis, logo existem vetores não-nulos (que podemos supor unitários) $u_1, u_2 \in E$ tais que $(A - \lambda_1 I)u_1 = 0$ e $(A - \lambda_2 I)u_2 = 0$, ou seja, $Au_1 = \lambda_1 u_1$ e $Au_2 = \lambda_2 u_2$. Pelo Teorema 13.4, $\{u_1, u_2\} \subset E$ é uma base ortonormal de autovetores de A. □

Corolário. *Todo operador auto-adjunto* $A\colon E \to E$*, num espaço vetorial de dimensão finita com produto interno, possui um autovetor.*

Com efeito, pelo Teorema 12.1 existe um subespaço $F \subset E$, de dimensão 1 ou 2, invariante por A. Se $\dim F = 1$ todo vetor não-nulo $v \in F$ é um autovetor de A. Se $\dim F = 2$ então, aplicando o Teorema 13.5 à restrição $A\colon F \to F$ de A ao subespaço invariante F, obtemos um autovetor $v \in F$. □

Teorema 13.6. (Teorema Espectral.) *Para todo operador auto-adjunto* $A\colon E \to E$*, num espaço vetorial de dimensão finita munido de produto interno, existe uma base ortonormal* $\{u_1, \ldots, u_n\} \subset E$ *formada por autovetores de* A.

Demonstração: Usaremos indução na dimensão de E. O teorema é evidente se $\dim E = 1$. Supondo-o verdadeiro em dimensão $n-1$, seja $\dim E = n$. Pelo Corolário do Teorema 13.5, existe um autovetor unitário u_n, portanto um subespaço $F \subset E$, de dimensão 1, invariante por A. Pelo Teorema 13.2, o complemento ortogonal $F^\perp$ também é invariante por A. Como $\dim F^\perp = n-1$, a hipótese de indução assegura a existência de uma base ortonormal $\{u_1, \ldots, u_{n-1}\} \subset F^\perp$ formada por autovetores da restrição $A\colon F^\perp \to F^\perp$. Segue-se que $\{u_1, \ldots, u_{n-1}, u_n\} \subset E$ é uma base ortonormal formada por autovetores de A. □

Observação. Vale a recíproca do Teorema Espectral: se existe uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$ formada por autovetores do operador $A \colon E \to E$ então este operador é auto-adjunto. Com efeito, para quaisquer $i, j = 1, \ldots, n$, tem-se

$$\langle Au_i, u_j \rangle = \langle \lambda_i u_i, u_j \rangle = \lambda_i \delta_{ij} = \lambda_j \delta_{ij} = \langle u_i, \lambda_j u_j \rangle = \langle u_i, Au_j \rangle$$

e daí resulta que $\langle Au, v \rangle = \langle u, Av \rangle$ para quaisquer $u, v \in E$.

Diremos que o operador linear $A \colon E \to E$ é *não-negativo*, e escreveremos $A \geq 0$, quando A for auto-adjunto e, além disso, $\langle Av, v \rangle \geq 0$ para todo $v \in E$. No caso particular em que $\langle Av, v \rangle > 0$ para todo $v \neq 0$, diremos que A é um operador *positivo* e escreveremos $A > 0$.

Teorema 13.7. *Um operador auto-adjunto $A \colon E \to E$ é não-negativo se, e somente se, seus autovalores são todos ≥ 0. A é positivo se, e somente se, todos os seus autovalores são números positivos.*

Demonstração: Se $A \geq 0$ e $Av = \lambda v$ com $v \neq 0$ então

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle Av, v \rangle \geq 0,$$

portanto $\lambda \geq 0$. Reciprocamente, se os autovalores de A são ≥ 0, seja $\{u_1, \ldots, u_n\} \subset E$ uma base ortonormal formada por autovetores (a qual existe pelo Teorema Espectral), com $Au_i = \lambda_i u_i$. Para todo vetor $v \in E$, tem-se $v = \alpha_1 u_1 + \cdots + \alpha_n u_n$, logo

$$\langle Av, v \rangle = \langle \Sigma \alpha_i A u_i, \Sigma \alpha_j u_j \rangle = \langle \Sigma \alpha_i \lambda_i u_i, \Sigma \alpha_j u_j \rangle = \Sigma \lambda_i \alpha_i^2.$$

Como $\lambda_i \geq 0$ para $i = 1, \ldots, n$, segue-se que $\langle Av, v \rangle \geq 0$, portanto $A \geq 0$. A afirmação sobre operadores positivos se prova da mesma maneira. □

Corolário 1. *Seja $A \geq 0$. Se, para um certo $v \in E$, vale $\langle Av, v \rangle = 0$ então $Av = 0$.*

Com efeito, sejam $\lambda_1, ..., \lambda_k$ os autovalores não-nulos de A. Então, pondo $v = \alpha_1 u_1 + \cdots + \alpha_n u_n$, resulta que $Av = \lambda_1 \alpha_1 u_1 + \cdots + \lambda_k \alpha_k u_k$, donde

$$0 = \langle Av, v \rangle = \lambda_1 \alpha_1^2 + \cdots + \lambda_k \alpha_k^2.$$

Como $\lambda_1 > 0, \ldots, \lambda_k > 0$, segue-se que $\alpha_1 = \cdots = \alpha_k = 0$, portanto $Av = 0$. □

Observação. Geometricamente, $A \geq 0$ significa que o ângulo entre v e Av (caso estes vetores sejam $\neq 0$) é sempre agudo ou reto. O Corolário 1 diz que, quando $Av \neq 0$, esse ângulo é sempre agudo.

Corolário 2. *Um operador é positivo se, e somente se, é não-negativo e invertível.*

Com efeito, se $A \geq 0$ é invertível então, para todo $v \neq 0$ tem-se $Av \neq 0$ logo, pelo Corolário 1, $\langle Av, v\rangle > 0$. A recíproca é óbvia. $\square$

Uma matriz quadrada $\mathbf{a} = [a_{ij}] \in M(n \times n)$ diz-se *não-negativa*, e escreve-se $\mathbf{a} \geq 0$, quando o operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$, cuja matriz na base canônica é $\mathbf{a}$, é não-negativo.

Dado $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$, tem-se $Av = (y_1, \ldots, y_n)$ onde, para cada $i = 1, \ldots, n$, $y_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$. Logo

$$\langle Av, v\rangle = \sum_{i=1}^{n} x_i y_i = \sum_{i,j=1}^{n} a_{ij} x_i x_j \,.$$

Portanto a matriz $\mathbf{a} = [a_{ij}] \in M(n \times n)$ é não-negativa se, e somente se, é simétrica e, para todo $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$ tem-se

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j \geq 0.$$

Analogamente, a matriz $\mathbf{a}$ diz-se *positiva* quando o operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$, que a ela corresponde, é positivo. Isto significa que $\mathbf{a}$ é simétrica e, para todo $v = (x_1, \ldots, x_n) \neq 0$ em $\mathbb{R}^n$, tem-se

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j > 0.$$

Assim uma matriz simétrica é não-negativa (respect. positiva) se, e somente se, seus autovalores são ≥ 0 (respect. positivos).

Exemplo 13.5. O polinômio característico da matriz simétrica $\mathbf{a} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ é $p(\lambda) = \lambda^2 - (a+c)\lambda + ac - b^2$. A soma de suas raízes é $a+c$ e o produto é $ac-b^2$. As raízes de $p(\lambda)$ são ambas positivas se, e somente se $ac-b^2 > 0$ e $a > 0$ (ou $c > 0$). Com efeito, $ac-b^2 > 0$ (isto é, $ac > b^2$) diz que essas raízes têm o mesmo sinal. De $a > 0$ tiramos

$c > b^2/a > 0$, logo $a + c > 0$ e o sinal comum das raízes é positivo. Portanto as condições $a > 0$ e $ac - b^2 > 0$ (ou $c > 0$ e $ac - b^2 > 0$) são necessárias e suficientes para que a matriz simétrica **a** seja positiva. Para que se tenha $\mathbf{a} \geq 0$ é necessário e suficiente que $ac - b^2 \geq 0$, $a \geq 0$ e $c \geq 0$. Assim, por exemplo, a matriz $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ é positiva e $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ é não-negativa.

Um operador $X \colon E \to E$ chama-se *raiz quadrada* do operador $A \colon E \to E$ quando $X^2 = A$.

Para uso na demonstração do Teorema 13.8, apresentaremos uma noção que tem outras aplicações.

Se λ é um autovalor do operador $A \colon E \to E$, o conjunto

$$E_\lambda = \{v \in E; Av = \lambda v\}$$

é um subespaço vetorial de E, invariante por A, chamado um *auto-subespaço*. Restrito a E_λ, o operador A é simplesmente a multiplicação por λ. Assim, todo vetor não-nulo em E_λ é um autovetor de A, com autovalor λ.

Quando A é auto-adjunto e $\lambda_1, \ldots, \lambda_r$ são seus autovalores *distintos* então, pelo Teorema 13.4, para $i \neq j$, todo vetor em E_{λ_i} é ortogonal a qualquer vetor em E_{λ_j}. Além disso, pelo Teorema Espectral, todo vetor $v \in E$ se escreve, de modo único, como $v = v_1 + \cdots + v_r$, onde $v_1 \in E_{\lambda_1}, \ldots, v_r \in E_{\lambda_r}$, ou seja, E é *soma direta* dos subespaços $E_{\lambda_1}, \ldots, E_{\lambda_r}$.

Teorema 13.8. *Todo operador não-negativo $A \colon E \to E$, num espaço vetorial de dimensão finita munido de produto interno, possui uma única raiz quadrada não-negativa, a qual é positiva se, e somente se, A é positivo.*

Demonstração: Sejam $\lambda_1 \geq 0, \ldots, \lambda_r \geq 0$ os autovalores distintos de A. Todo vetor $v \in E$ se escreve, de modo único, como $v = v_1 + \cdots + v_r$, onde $v_1 \in E_{\lambda_1}, \ldots, v_r \in E_{\lambda_r}$, logo $Av = \lambda_1 v_1 + \cdots + \lambda_r v_r$. Definimos o operador linear $B \colon E \to E$ pondo

$$Bv = \sqrt{\lambda_1} \cdot v_1 + \cdots + \sqrt{\lambda_r} \cdot v_r.$$

Se $w = w_1 + \cdots + w_r$ com $w_1 \in E_{\lambda_1}, \ldots, w_r \in E_{\lambda_r}$ então $\langle v_i, w_j \rangle = 0$

para $i \neq j$, portanto

$$\langle Bv, w\rangle = \sum_{i=1}^{r} \sqrt{\lambda_i}\,\langle v_i, w_i\rangle = \langle v, Bw\rangle \quad \text{e} \quad \langle Bv, v\rangle = \sum_{i=1}^{r} \sqrt{\lambda_i}|v_i|^2.$$

Logo B é um operador não-negativo. Além disso, é claro que $B^2v = Av$ para todo $v \in E$, logo B é uma raiz quadrada de A. Para provar que é única a raiz quadrada não-negativa de A, começamos observando que toda raiz quadrada de A comuta com A. Com efeito, se $C^2 = A$ então $AC = C^2C = CC^2 = CA$. A observação seguinte é que se C comuta com A então cada um dos auto-subespaços de A é invariante por C. Com efeito,

$$v \in E_{\lambda_i} \Rightarrow Av = \lambda_i v \Rightarrow A(Cv) = C(Av) = C(\lambda_i v) = \lambda_i(Cv),$$

logo $Cv \in E_{\lambda_i}$.

Mais uma observação: se C é uma raiz quadrada não-negativa de A então, para cada $i = 1, \ldots, r$, o único autovalor da restrição de C a E_{λ_i} é $\sqrt{\lambda_i}$. Com efeito, se $w \in E_{\lambda_i}$ é autovetor de C, digamos com $Cw = \gamma \cdot w$, então $\lambda_i w = Aw = C^2w = \gamma^2 \cdot w$, logo $\lambda_i = \gamma^2$ e $\gamma = \sqrt{\lambda_i}$. Agora o argumento final: seja C uma raiz quadrada não-negativa de A. Cada auto-subespaço E_{λ_i} é invariante pelo operador auto-adjunto C e o único autovalor de C em E_{λ_i} é $\sqrt{\lambda_i}$. Logo $Cw = \sqrt{\lambda_i} \cdot w$ para todo $w \in E_{\lambda_i}$. Então, dado $v = v_1 + \cdots + v_r$, com $v_1 \in E_{\lambda_1}, \ldots, v_r \in E_{\lambda_r}$, tem-se

$$Cv = \sqrt{\lambda_1} \cdot v_1 + \cdots + \sqrt{\lambda_r} \cdot v_r$$

portanto C coincide com o operador B acima definido. A afirmação sobre a positividade da raiz quadrada é óbvia. □

Observação 1: Um dos argumentos acima usados permite mostrar que se dois operadores auto-adjuntos A, B comutam então eles têm um autovetor comum. Com efeito, seja λ_1 um autovalor de A. Como vimos acima, a hipótese $AB = BA$ implica que o subespaço $E_{\lambda_1} = \{v \in E; Av = \lambda_1 v\}$ é invariante por B. Pelo corolário do Teorema 13.5, o operador auto-adjunto B possui um autovetor $w \in E_{\lambda_1}$. Mas todo vetor não-nulo em E_{λ_1} é autovetor de A. Logo w é autovetor comum de A e B. Usando repetidamente o Teorema 13.2 conclui-se então que se dois operadores auto-adjuntos comutam então existe uma base ortonormal formada por autovetores de ambos.

Observação 2: Somente operadores não-negativos possuem raiz quadrada auto-adjunta, pois resulta imediatamente da definição de A^* que o quadrado de um operador auto-adjunto é não-negativo. Entretanto, o quadrado de um operador de outro tipo pode ser um operador (auto-adjunto) negativo. Por exemplo, a rotação de $90°$ no plano tem quadrado igual à rotação de $180°$, que é igual a $-I$. Além disso, um operador positivo pode ter uma raiz quadrada que não é auto-adjunta. Por exemplo, o operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, definido por $A(x,y) = (2x-y, 3x-2y)$ é uma raiz quadrada da identade $I_{\mathbb{R}^2}$, como se pode ver sem dificuldade.

Exemplos gerais de operadores não-negativos são dados pelo teorema seguinte.

Teorema 13.9. *Seja $A\colon E \to F$ uma transformação linear entre espaços vetoriais de dimensão finita munidos de produto interno. Os operadores $A^*A\colon E \to E$ e $AA^*\colon F \to F$ são não-negativos e têm ambos o mesmo posto de A (e de A^*). Em particular, são positivos se, e somente se, A é invertível.*

Demonstração: Como $(A^*A)^* = A^*A^{**} = A^*A$, vemos que A^*A é auto-adjunto e, semelhantemente, AA^* também. Além disso, para todo $v \in E$, tem-se $\langle A^*Av, v\rangle = \langle Av, Av\rangle = |Av|^2 \geq 0$ logo $A^*A \geq 0$. Da mesma forma se vê que $AA^* \geq 0$. Para determinar o posto de A^*A, mostraremos inicialmente que $\mathcal{N}(A^*A) = \mathcal{N}(A)$. A inclusão $\mathcal{N}(A) \subset \mathcal{N}(A^*A)$ é óbvia, pois $\mathcal{N}(A) \subset \mathcal{N}(BA)$ seja qual for $B\colon F \to E$. Por outro lado,

$$\begin{aligned} v \in \mathcal{N}(A^*A) &\Rightarrow A^*Av = 0 \\ &\Rightarrow Av \in \mathcal{N}(A^*) = \mathcal{I}m(A)^{\perp} \\ &\Rightarrow Av \in \mathcal{I}m(A) \cap \mathcal{I}m(A)^{\perp}, \end{aligned}$$

logo $v \in \mathcal{N}(A^*A) \Rightarrow Av = 0$, ou seja $\mathcal{N}(A^*A) \subset \mathcal{N}(A)$, que é a inclusão restante. Em seguida observemos que, pelo Teorema do Núcleo e da Imagem:

$$\begin{aligned} \text{posto de } A^*A &= \dim E - \dim \mathcal{N}(A^*A) \\ &= \dim E - \dim \mathcal{N}(A) \\ &= \text{ posto de } A. \end{aligned}$$

A afirmação análoga sobre AA^* se prova do mesmo modo. □

Corolário. *A transformação linear* $A\colon E \to F$ *é injetiva se, e somente se,* A^*A *é invertível.* A *é sobrejetiva se, e somente se,* AA^* *é invertível.*

Com efeito, A injetiva $\Leftrightarrow$ posto de $A = \dim E \Leftrightarrow$ posto de $A^*A = \dim E \Leftrightarrow A^*A$ invertível. Mesmo argumento para AA^*. □

Observação: O Teorema 13.9 e seu corolário podem ser enunciados em termos de matrizes, substituindo-se A por $\mathbf{a}$ e A^* por $\mathbf{a}^T$. Nesta formulação, dizer que A é injetiva equivale a afirmar que o posto da matriz $\mathbf{a} \in M(m \times n)$ é igual a n. Analogamente, A ser sobrejetiva equivale à matriz $\mathbf{a}$ ter posto m. Qualquer destas duas hipóteses sobre A, quando formulada em termos de matrizes, pode ser expressa na afirmação única de que $\mathbf{a}$ tem *posto máximo*, isto é, o posto de $\mathbf{a} \in M(m \times n)$ é o menor dos dois números m e n. Na prática, quando quisermos um exemplo de matriz positiva ou não-negativa, tomamos uma matriz $\mathbf{a}$ de posto máximo e consideramos as matrizes simétricas $\mathbf{aa}^T$ e $\mathbf{a}^T\mathbf{a}$. A maior delas é ≥ 0 e a menor é > 0.

Exemplo 13.6. Seja $\mathbf{a} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. Então as matrizes

$$\mathbf{aa}^T = \begin{bmatrix} 14 & 32 \\ 32 & 77 \end{bmatrix} \quad \text{e} \quad \mathbf{a}^T\mathbf{a} = \begin{bmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{bmatrix},$$

ambas de posto 2, são positiva e não-negativa, respectivamente.

A seguir, uma extensão do Teorema Espectral, válida para transformações lineares quaisquer.

Teorema 13.10. (Teorema dos Valores Singulares.) *Seja* $A\colon E \to F$ *uma transformação linear de posto* r *entre espaços de dimensão finita com produto interno. Existem bases ortonormais* $\{u_1, \ldots, u_n\} \subset E$ *e* $\{v_1, \ldots, v_m\} \subset F$ *tais que* $Au_i = \sigma_i v_i$, $A^*v_i = \sigma_i u_i$, *com* $\sigma_i > 0$ *para* $i = 1, \ldots, r$ *e* $Au_i = 0$, $A^*v_i = 0$ *se* $i > r$.

Demonstração: Pelo Teorema 13.9, o operador $A^*A\colon E \to E$ é não-negativo e tem posto r. O Teorema Espectral assegura a existência de uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$ tal que $A^*Au_i = \sigma_i^2 u_i$, com $\sigma_i > 0$ se $1 \leq i \leq r$ e $\sigma_i = 0$ se $r + 1 \leq i \leq n$. Então

$$\langle Au_i, Au_j \rangle = \langle u_i, A^*Au_j \rangle = \sigma_j^2 \langle u_i, u_j \rangle.$$

Assim, os vetores $Au_1, \ldots, Au_r$ são dois a dois ortogonais, e não-nulos pois

$$|Au_i|^2 = \sigma_i^2 |u_i|^2,$$

logo $|Au_i| = \sigma_i$. (Na realidade, a prova do Teorema 13.9 mostra que $\mathcal{N}(A) = \mathcal{N}(A^*A)$, portanto $Au_j = 0$ se $r + 1 \leq j \leq n$.) Podemos então escrever, para $i = 1, \ldots, r$, $Au_i = \sigma_i v_i$, onde $\{v_1, \ldots, v_r\} \subset F$ é um conjunto ortonormal, de fato uma base ortonormal de $\mathcal{I}m(A)$, a qual pode ser completada a uma base ortonormal $\{v_1, \ldots, v_m\} \subset F$, onde $\{v_{r+1}, \ldots, v_m\}$ é uma base ortonormal de $\mathcal{N}(A^*)$. Tem-se $A^*v_i = 0$ se $i > r$ e, para $i = 1, \ldots, r$:

$$A^*v_i = \frac{1}{\sigma_i} A^*Au_i = \frac{1}{\sigma_i} \sigma_i^2 u_i = \sigma_i u_i. \quad \square$$

Os números positivos $\sigma_1, \ldots, \sigma_r$ chamam-se os *valores singulares* da transformação linear $A \colon E \to F$, de posto r.

No teorema acima, podemos observar que $\{v_1, \ldots, v_m\} \subset F$ é uma base ortonormal de autovetores para o operador $AA^* \colon F \to F$, pois

$$(AA^*)v_i = A(\sigma_i u_i) = \sigma_i Au_i = \sigma_i^2 v_i.$$

Analogamente, $\{u_1, \ldots, u_n\} \subset E$ é uma base ortonormal de autovetores do operador $A^*A \colon E \to E$.

Vemos ainda que $\{u_1, \ldots, u_r\}$ é uma base ortonormal de $\mathcal{I}m(A^*)$, enquanto o conjunto ortonormal $\{v_1, \ldots, v_r\}$ é uma base para $\mathcal{I}m(A)$. Ao mesmo tempo, $\{u_{r+1}, \ldots, u_n\} \subset \mathcal{N}(A)$ e $\{v_{r+1}, \ldots, v_m\} \subset \mathcal{N}(A^*)$ são bases, desde que não sejam vazios.

Exercícios

Em todos os exercícios desta seção E é um espaço vetorial de dimensão finita, munido de produto interno.

13.1. Prove que duas projeções $P, Q \colon E \to E$ são iguais se, e somente se, têm os mesmos auto-vetores com os mesmos auto-valores. Conclua que uma projeção $P \colon E \to E$ é auto-adjunta (portanto $\mathcal{I}m(P) = \mathcal{N}(P)^\perp$) se, e somente se, é normal. (Veja Exerc. 12.7.)

13.2. Sejam $A, B \colon E \to E$ operadores auto-adjuntos tais que $\langle Av, v \rangle = \langle Bv, v \rangle$ para todo $v \in E$. Prove que $A = B$.

13.3. Se B é invertível e BAB^* é auto-adjunto, prove que A é auto-adjunto.

13.4. Sejam P uma projeção ortogonal e $\alpha > 0$. Exprima a raiz quadrada positiva de $I + \alpha P$ em termos de P.

13.5. Seja A auto-adjunto. Prove que $A^k v = 0$ implica $Av = 0$.

13.6. Assinale se cada um dos seguintes subconjuntos do espaço vetorial $\mathcal{L}(E)$ é um subespaço vetorial (S), um cone (C), um cone convexo (CC):

( ) operadores normais (Veja Exerc. 12.7.)

( ) operadores auto-adjuntos

( ) operadores não-negativos

( ) homotetias.

13.7. Sejam $S, T \in \mathcal{L}(E)$ involuções auto-adjuntas. Prove que ST é uma involução auto-adjunta se, e somente se, $ST = TS$.

13.8. Dados os vetores $v = (2, -1, -2)$, e $w = (3, -6, -6)$, determine o operador auto-adjunto $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ tal que $Av = (1, 1, 13)$ e $Aw = (3, 21, 33)$, sabendo que o traço de A é 5.

13.9. Dados os vetores $u = (4, 4, -2)$, $v = (4, -2, 4)$ e $w = (1, -2, -2)$, seja $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ o operador linear tal que $Au = (10, -2, -2)$, $Av = (-2, 10, -2)$ e $Aw = (1, 1, -5)$. Prove que A é auto-adjunto.

13.10. Dado o subespaço $F \subset E$, considere as transformações lineares $P\colon E \to F$ e $J\colon F \to E$, onde P é a projeção ortogonal (de núcleo $F^\perp$) e $Jv = v$ para todo $v \in F$. (J chama-se a *inclusão* de F em E.) Determine as adjuntas $P^*\colon F \to E$ e $J^*\colon E \to F$.

13.11. Seja $A\colon E \to E$ o operador de posto 1 definido por $Av = \langle v, a \rangle\, b$. ($E$ é um espaço vetorial de dimensão finita, com produto interno, e $a, b \in E$ são vetores não-nulos.) Prove que A é auto-adjunto se, e somente se, b é múltiplo de a. Além disso, A é não-negativo se, e somente se, pode-se tomar $b = a$. Dada uma base ortonormal em E, determine a matriz de A em função das coordenadas de a e b nessa base. (Veja Exercício 10.32.)

13.12. Se $A^*A = -A$, prove que os autovalores de A pertencem ao conjunto $\{0, -1\}$. Dê exemplo de uma matriz $\mathbf{a} \in M(2 \times 2)$, tal que $a_{11} = -\frac{1}{3}$ e $\mathbf{a}^T\mathbf{a} = -\mathbf{a}$. Quantas dessas matrizes existem?

13.13. Seja $A: E \to E$ auto-adjunto. Para todo $k \in \mathbb{N}$ ímpar, mostre que existe um único operador auto-adjunto $X: E \to E$ tal que $X^k = A$. Se k é par, existe X auto-adjunto com $X^k = A$ se, e somente se, $A \geq 0$. Neste caso, X pode ser escolhido ≥ 0 e então é único.

13.14. Assinale V(erdadeiro) ou F(also):

( ) Se todos os elementos de uma matriz simétrica são números positivos então essa matriz é não-negativa.

( ) O produto de operadores não-negativos é um operador não-negativo.

( ) Um operador não-negativo é positivo ou é zero.

( ) Uma matriz do tipo $[a_i \cdot b_j]$ é não-negativa se, e somente se, $a_i = b_i$ (para todo i).

( ) O posto de uma matriz não-negativa $n \times n$ pode ser qualquer número de 1 a n.

( ) O inverso de um operador auto-adjunto (invertível) também é auto-adjunto.

( ) Se existirem $u, v \in \mathbb{R}^3$ não-nulos, com $Au = 2u$, $Av = 3v$ então existe uma base de autovetores para o operador linear $A: \mathbb{R}^3 \to \mathbb{R}^3$.

( ) Sejam $\langle\ ,\ \rangle$ e $[\ ,\ ]$ produtos internos definidos em $\mathbb{R}^2$. Se um operador A é auto-adjunto relativamente a $\langle\ ,\ \rangle$ então A também é auto-adjunto em relação a $[\ ,\ ]$.

13.15. Se o espaço vetorial E possui uma base formada por autovetores do operador $A: E \to E$, prove que é possível definir em E um produto interno em relação ao qual A é auto-adjunto.

13.16. Num espaço vetorial E, de dimensão finita, seja A um operador diagonalizável (ou seja, E possui uma base formada por autovetores de A). Se $F \subset E$ é um subespaço invariante por A, prove que

a restrição de A ao subespaço F é um operador diagonalizável em F. (Sugestão: use o exercício anterior.)

13.17. Seja $A \colon E \to E$ um operador diagonalizável. Se o subespaço $F_1 \subset E$ é invariante por A, prove que existe um subespaço $F_2 \subset E$, também invariante por A, tal que $E = F_1 \oplus F_2$.

13.18. Se os operadores $A, B \colon E \to E$ são auto-adjuntos, prove que $AB + BA$ é auto-adjunto. Que se pode dizer sobre $AB - BA$?

13.19. Se $A \colon E \to E$ é auto-adjunto, prove que, para todo $B \in \mathcal{L}(E)$, o operador B^*AB também é auto-adjunto. Se $A \geq 0$, prove que $B^*AB \geq 0$. Se $A > 0$ e B é invertível, prove que $B^*AB > 0$.

13.20. Se dois operadores auto-adjuntos $A, B \colon E \to E$ comutam, prove que o espaço E possui uma base ortonormal formada por auto-vetores comuns a A e B. Prove também a recíproca.

13.21. Assinale V(erdadeiro) ou F(also):

( ) O conjunto dos operadores positivos é um cone convexo no espaço vetorial $\mathcal{L}(E)$.

( ) O conjunto dos operadores não-negativos é um subespaço vetorial de $\mathcal{L}(E)$.

( ) Os elementos da diagonal de uma matriz positiva são números positivos.

( ) Se A é auto-adjunto e B é invertível então $B^{-1}AB$ é auto-adjunto.

( ) Existe uma matriz positiva 2×2 com dois elementos negativos e dois positivos.

( ) Se $A \colon \mathbb{R}^n \to \mathbb{R}^n$ é um operador invertível qualquer, alguma base ortogonal de $\mathbb{R}^n$ é transformada por A numa base ortogonal.

( ) Se o operador A é auto-adjunto então o traço de A é igual à soma dos seus autovalores.

13.22. Seja E um espaço vetorial de dimensão finita, com produto interno. Um subconjunto $\Sigma \subset E$ chama-se um *elipsóide* quando

existem uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$ e números positivos $a_1, \ldots, a_n$ tais que Σ é o conjunto dos vetores $v = x_1u_1 + \cdots + x_nu_n$ cujas coordenadas x_i satisfazem a equação $a_1x_1^2 + \cdots + a_nx_n^2 = 1$. Seja $A\colon E \to E$ um operador invertível. Prove que todo elipsóide Σ é transformado por A num elipsóide Σ'.
(Sugestão: use o Teorema 13.10.)

13.23. Seja Σ um subconjunto de um espaço vetorial de dimensão finita, com produto interno. Prove que Σ é um elipsóide se, e somente se, existe um operador positivo $A\colon E \to E$ tal que

$$\Sigma = \{v \in E; \langle Av, v\rangle = 1\}.$$

13.24. Num espaço vetorial E, de dimensão finita, com produto interno $\langle u, v\rangle$, seja B um operador positivo. Prove que $[u, v] = \langle Bu, v\rangle$ define um novo produto interno em E. Se $A\colon E \to E$ é auto-adjunto no sentido do produto interno original, prove que A é também auto-adjunto no sentido do novo produto interno se, e somente se, $AB = BA$.

13.25. Sejam $A\colon E \to E$ auto-adjunto e $B\colon E \to E$ positivo. Prove:

(a) Se X é a raiz quadrada positiva de B então XAX é auto-adjunto.

(b) v é autovetor de XAX se, e somente se, Xv é autovetor de BA.

(c) E possui uma base (não necessariamente ortogonal) de autovetores de BA. (Ou seja, BA é diagonalizável.)

13.26. Se $A\colon E \to E$ é auto-adjunto e $B\colon E \to E$ é positivo, prove que E possui uma base $\mathcal{V}$ tal que, para todo $v \in \mathcal{V}$, existe $\lambda \in \mathbb{R}$ com $Av = \lambda Bv$. ("Problema de autovalores generalizados".)

13.27. Sejam A, B operadores auto-adjuntos no mesmo espaço vetorial. Se BA é diagonalizável, prove que AB também é diagonalizável. (Veja Exercício 12.35.)

13.28. Dada a transformação linear $A\colon E \to F$, entre espaços vetoriais de dimensão finita munidos de produto interno, seja σ^2 o maior autovalor do operador $A^*A\colon E \to E$ $(\sigma \geq 0)$. Prove que σ é o maior dos números $|Au|$, onde u é qualquer vetor unitário em E. Escreve-se $||A|| = \sigma = \max\{|Au|; u \in E, |u| = 1\}$ e diz-se que $||A||$ é a *norma*

espectral da transformação linear A. Seja $|A| = \sqrt{\mathrm{tr}(A^*A)}$ a norma induzida pelo produto interno $\langle A, B\rangle = \mathrm{tr}(A^*B)$, definido no Exercício 11.17. Se $\sigma_1^2, \ldots, \sigma_n^2$ são os autovalores de A^*A, tem-se

$$|A| = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2} \quad \text{e} \quad \|A\| = \max \sigma_i.$$

Conclua que $\|A\| \leq |A| \leq \sqrt{n}.\|A\|$.

13.29. Prove que todo operador auto-adjunto $A\colon E \to E$ pode ser escrito como $A = \lambda_1 P_1 + \cdots + \lambda_m P_m$ onde:

(1) $\lambda_1 < \cdots < \lambda_m$.

(2) $P_1^2 = P_1 = P_1^*, \ldots, P_m^2 = P_m = P_m^*$. (Cada P_i é uma projeção ortogonal.)

(3) $P_i P_j = 0$ se $i \neq j$.

(4) $P_1 + \cdots + P_m = I$.

Prove também que a expressão $A = \Sigma\lambda_i P_i$ com as propriedades (1) a (4) acima é única. (Sugestão: $\lambda_1 < \cdots < \lambda_m$ são os autovalores distintos de A e P_i é a projeção ortogonal sobre o auto-espaço E_{λ_i}, definido no Exercício 12.23.)

13.30. Prove que todo operador auto-adjunto é soma de operadores auto-adjuntos de posto 1, os quais podem ser tomados não-negativos se A for não-negativo. [Sugestão: Teorema Espectral.]

13.31. Chama-se *produto de Hadamard* de duas matrizes $\mathbf{a} = [a_{ij}]$, $\mathbf{b} = [b_{ij}] \in M(m \times n)$ à matriz $\mathbf{c} = [c_{ij}] \in M(m \times n)$ com $c_{ij} = a_{ij} \cdot b_{ij}$. Prove que o produto de Hadamard de duas matrizes não-negativas é uma matriz não-negativa. (Use o fato de que toda matriz não-negativa pode escrever-se como soma

$$\mathbf{a} = \sum_r \mathbf{a}^{(r)}$$

de matrizes não-negativas de posto 1, cada uma das quais tem a forma $\mathbf{a}^{(r)} = [a_i^{(r)} \cdot a_j^{(r)}]$.) (Veja o exercício anterior.)

13.32. Prove que a norma espectral, definida no Exercício 13.18, tem as propriedades

$$\|A + B\| \leq \|A\| + \|B\| \quad \text{e} \quad \|BA\| \leq \|B\|\,\|A\|.$$

Dê exemplo de um operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ para o qual se tem $\|A^2\| < \|A\|^2$.

13.33. Prove que a norma $|A| = \sqrt{\operatorname{tr}(A^*A)}$, proveniente do produto interno introduzido no Exercício 11.17, cumpre a desigualdade $|BA| \leq |B|\,|A|$.

13.34. Seja $A\colon E \to E$ auto-adjunto. Se $u \in E$ é tal que $Au \neq 0$ e $\langle Au, u\rangle = 0$, mostre que existem $v, w \in E$ com $\langle Av, v\rangle > 0$ e $\langle Aw, w\rangle < 0$. Deduza daí o Cor. 1 do Teor. 13.7 sem usar o Teorema Espectral. (Sugestão: considere $tu + Au$ para valores convenientes de t.)

14

Operadores Ortogonais

Sob o ponto de vista da organização geral da Matemática, os operadores ortogonais são os automorfismos da estrutura de espaço vetorial com produto interno, ou seja, são as simetrias dessa estrutura. Do ponto de vista mais pedestre em que nos colocamos, os operadores ortogonais são aqueles para os quais se podem obter as matrizes mais simples, depois dos auto-adjuntos. Eles possuem as propriedades geométricas mais marcantes, muitas das quais lhes são exclusivas. (Vide Teorema 14.1.) Nesta seção consideramos, mais geralmente, as transformações lineares ortogonais de um espaço noutro e as matrizes ortogonais não-quadradas. Obtemos a forma mais simples que pode assumir a matriz de um operador ortogonal e concluímos demonstrando que todo operador é o produto de um não- negativo por um ortogonal (forma polar).

Nesta seção, estenderemos para espaços vetoriais com produto interno a noção de congruência entre figuras da Geometria Elementar. Lembramos que uma congruência entre duas figuras X, Y é uma bijeção $f\colon X \to Y$ que preserva distâncias, isto é, tal que $d(f(x), f(x')) = d(x, x')$ para quaisquer $x, x' \in X$. Note-se que, embora não garanta a sobrejetividade, a propriedade de preservar distâncias já assegura que f é injetiva pois $f(x) = f(x') \Rightarrow d(x, x') = d(f(x), f(x')) = 0 \Rightarrow x = x'$.

Na discussão a seguir, um papel central é desempenhado pelos conjuntos ortonormais $\{u_1, \ldots, u_n\} \subset \mathbb{R}^m$. Uma matriz $\mathbf{u} \in M(m \times n)$ cujas n colunas formam um conjunto ortonormal em $\mathbb{R}^m$ chama-se uma *matriz ortogonal*.

Se $u_1, \ldots, u_n \in \mathbb{R}^m$ são as colunas da matriz $\mathbf{u} = [a_{ij}] \in M(m \times n)$, a condição para que $\mathbf{u}$ seja ortogonal é que $\langle u_i, u_j \rangle = 0$ se $i \neq j$ e $\langle u_i, u_i \rangle = 1$, onde $i, j = 1, \ldots, n$. Noutras palavras, deve-se ter

$$\sum_{k=1}^{m} a_{ki} a_{kj} = \delta_{ij}. \tag{*}$$

(Estamos usando aqui, mais uma vez, o símbolo δ_{ij}, conhecido como "delta de Kronecker", que vale 0 se $i \neq j$ e 1 se $i = j$.) A igualdade (*) significa precisamente que o produto $\mathbf{u}^T \cdot \mathbf{u}$ da transposta $\mathbf{u}^T$ por $\mathbf{u}$ é igual à matriz identidade $n \times n$.

Portanto a matriz $\mathbf{u} \in M(m \times n)$ é ortogonal se, e somente se, $\mathbf{u}^T \cdot \mathbf{u} = \mathbf{I}_n$.

Se $\mathbf{u} \in M(m \times n)$ é ortogonal então seu posto é n (logo $m \geq n$), pois suas n colunas são vetores L.I. no espaço $\mathbb{R}^m$. Quando $m = n$ e $\mathbf{u} \in M(n \times n)$ é uma matriz quadrada ortogonal então a igualdade $\mathbf{u}^T \cdot \mathbf{u} = \mathbf{I}_n$ implica $\mathbf{u} \cdot \mathbf{u}^T = \mathbf{I}_n$ logo $\mathbf{u}^T = \mathbf{u}^{-1}$. (Vide Corolário do Teorema 6.7 ou a propriedade 7) do produto de matrizes na Seção 8.) Assim, matrizes quadradas ortogonais são aquelas cuja transposta é igual à inversa.

A igualdade $\mathbf{u} \cdot \mathbf{u}^T = \mathbf{I}_n$ significa que as linhas de $\mathbf{u}$ formam um conjunto de n vetores ortonormais em $\mathbb{R}^m$. Portanto, para matrizes quadradas, colunas ortonormais equivale a linhas ortonormais.

Se $\mathbf{u} \in M(n \times p)$, $\mathbf{v} \in M(m \times n)$ são ortogonais então $(\mathbf{vu})^T \mathbf{vu} = \mathbf{u}^T \mathbf{v}^T \mathbf{vu} = \mathbf{u}^T \mathbf{I}_n \mathbf{u} = \mathbf{u}^T \mathbf{u} = \mathbf{I}_p$ logo o produto $\mathbf{vu}$ é ortogonal. Se $\mathbf{u} \in M(n \times n)$ é ortogonal então $(\mathbf{u}^T)^T \cdot \mathbf{u}^T = \mathbf{u} \cdot \mathbf{u}^T = \mathbf{I}_n$ logo $\mathbf{u}^T = \mathbf{u}^{-1}$ também é ortogonal.

Evidentemente, se $m > n$, as m linhas de uma matriz ortogonal $\mathbf{u} \in M(m \times n)$ não podem formar um conjunto ortonormal pois são vetores em $\mathbb{R}^n$, logo $\mathbf{u}^T$ não é ortogonal quando $\mathbf{u}$ não é quadrada.

Exemplo 14.1. A matriz $\mathbf{u}$, cujas colunas são os vetores

$$u_1 = \left(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right),$$
$$u_2 = \left(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}\right) \quad \text{e}$$
$$u_3 = \left(\frac{2}{3}, \frac{2}{3}, -\frac{1}{3}\right),$$

é ortogonal.

Exemplo 14.2. (Matrizes ortogonais 2×2.) Seja

$$\mathbf{u} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

uma matriz ortogonal 2×2. Como $u_1 = (a, c)$ é um vetor unitário em $\mathbb{R}^2$, existe $\theta \in \mathbb{R}$ tal que $a = \cos\theta$, $c = \operatorname{sen}\theta$. Sendo $u_2 = (b, d)$ unitário e perpendicular a u_1, devemos ter $u_2 = \pm(-\operatorname{sen}\theta, \cos\theta)$. Assim, há duas possibilidades para a matriz $\mathbf{u}$:

$$\mathbf{u} = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \qquad \text{ou} \qquad \mathbf{u} = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{bmatrix}.$$

No primeiro caso, tem-se $ad - bc = 1$ e no segundo $ad - bc = -1$.

No primeiro caso, o polinômio característico de $\mathbf{u}$, $p(\lambda) = \lambda^2 - (2\cos\theta)\lambda + 1$, não tem raízes reais salvo se $\theta = 0$ ou $\theta = 180°$, casos em que $\mathbf{u} = \pm\mathbf{I}_2$. Trata-se da matriz de uma rotação. No segundo caso, $p(\lambda) = \lambda^2 - 1$ tem raízes ± 1. Então o operador $U\colon \mathbb{R}^2 \to \mathbb{R}^2$ cuja matriz (na base canônica) é $\mathbf{u}$ admite autovetores v_1, v_2, com $Uv_1 = v_1$ e $Uv_2 = -v_2$. Observe que, neste caso, a matriz $\mathbf{u}$ é simétrica, o operador U é auto-adjunto, $\{v_1, v_2\} \subset \mathbb{R}^2$ é uma base ortonormal e U é a reflexão em torno do eixo que contém v_1, paralelamente a v_2.

Exemplo 14.3. (Matriz de passagem ortogonal.) Sejam $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ e $\mathcal{U}' = \{u'_1, \ldots, u'_n\} \subset E$ bases ortonormais. A matriz de passagem $\mathbf{p} = [p_{ij}]$ de $\mathcal{U}$ para $\mathcal{U}'$ é uma matriz ortogonal. Com efeito, para quaisquer $i, j = 1, \ldots, n$, temos

$$u'_i = \sum_{k=1}^{n} p_{ki} u_k \quad \text{e} \quad u'_j = \sum_{k=1}^{n} p_{kj} u_k,$$

logo

$$\delta_{ij} = \langle u'_i, u'_j \rangle = \sum_{k=1}^{n} p_{ki} p_{kj} \, ,$$

portanto $\mathbf{p}^T \cdot \mathbf{p} = \mathbf{I}_n$. As últimas 2 igualdades da linha acima mostram que, reciprocamente, se a matriz de passagem $\mathbf{p}$ é ortogonal, então $\mathcal{U}$ ortonormal $\Rightarrow \mathcal{U}'$ ortonormal.

Teorema 14.1. *As seguintes afirmações a respeito de uma transformação linear* $A \colon E \to F$*, entre espaços vetoriais de dimensão finita providos de produto interno, são equivalentes:*

(1) A *preserva norma:* $|Av| = |v|$ *para todo* $v \in E$*;*

(2) A *preserva distância:* $|Au - Av| = |u - v|$ *para quaisquer* $u, v \in E$*;*

(3) A *preserva produto interno:* $\langle Au, Av \rangle = \langle u, v \rangle$ *para quaisquer* $u, v \in E$*;*

(4) $A^*A = I_E$ *;*

(5) *A matriz de* A *relativa a qualquer par de bases ortonormais* $\mathcal{U} \subset E$*,* $\mathcal{V} \subset F$ *é uma matriz ortogonal;*

(6) *A matriz de* A *relativa a um certo par de bases ortonormais* $\mathcal{U} \subset E$*,* $\mathcal{V} \subset F$ *é uma matriz ortogonal.*

(7) A *transforma uma certa base ortonormal* $\mathcal{U} \subset E$ *num conjunto ortonormal* $\mathcal{X} \subset F$*. (Se* $\dim E = \dim F$*,* $\mathcal{X}$ *é uma base.)*

(8) A *transforma toda base ortonormal* $\mathcal{W} \subset E$ *num conjunto ortonormal* $\mathcal{Z} \subset F$*.*

Demonstração: Se vale (1) então $|Au - Av| = |A(u - v)| = |u - v|$, logo $(1) \Rightarrow (2)$. Se vale (2) então

$$\begin{aligned} \langle Au, Av \rangle &= \frac{1}{2}\left[|Au|^2 + |Av|^2 - |Au - Av|^2\right] \\ &= \frac{1}{2}\left[|u|^2 + |v|^2 - |u - v|^2\right] = \langle u, v \rangle \, , \end{aligned}$$

logo $(2) \Rightarrow (3)$. Se vale (3) então, para quaisquer $u, v \in E$ tem-se $\langle u, v \rangle = \langle Au, Av \rangle = \langle A^*Au, v \rangle$, portanto $A^*Au = u$ para todo $u \in E$,

donde $A^*A = I_E$, logo $(3) \Rightarrow (4)$. Se vale (4) e $\mathbf{a}$ é a matriz de A nas bases ortonormais $\mathcal{U} \subset E$, $\mathcal{V} \subset F$ então $\mathbf{a}^T \cdot \mathbf{a} = \mathbf{I}_n$ e $\mathbf{a}$ é uma matriz ortogonal, logo $(4) \Rightarrow (5)$. Obviamente $(5) \Rightarrow (6)$. Se vale (6), sejam $\mathcal{U} = \{u_1, \ldots, u_n\}$, $\mathcal{V} = \{v_1, \ldots, v_m\}$ e $\mathbf{a} = [a_{ij}]$ a matriz ortogonal de A nessas bases. De

$$Au_i = \sum_{k=1}^{m} a_{ki} v_k \quad \text{e} \quad Au_j = \sum_{k=1}^{m} a_{kj} v_k$$

resulta

$$\langle Au_i, Au_j \rangle = \sum_{k=1}^{m} a_{ki} a_{kj} = \delta_{ij},$$

logo $\mathcal{X} = \{x_1, \ldots, x_n\} \subset F$, com $x_i = Au_i$, é um conjunto ortonormal e $(6) \Rightarrow (7)$. Se vale (7), seja $\mathcal{W} = \{w_1, \ldots, w_n\} \subset E$ uma base ortonormal qualquer. Para $i, j = 1, \ldots, n$, temos

$$w_i = \sum_{k} p_{ki} u_k \quad \text{e} \quad w_j = \sum_{k} p_{kj} u_k,$$

onde a matriz de passagem $\mathbf{p} = (p_{ij})$ é ortogonal, pelo Exemplo 14.3. Pondo $\mathcal{Z} = \{z_1, \ldots, z_n\}$, onde $z_i = Aw_i$, vem, para $i, j = 1, \ldots, n$:

$$z_i = \sum_{k=1}^{n} p_{ki} A u_k = \sum_{k=1}^{n} p_{ki} x_k \quad \text{e} \quad z_j = \sum_{k=1}^{n} p_{kj} x_k\,.$$

Como $\mathcal{X} = \{x_1, \ldots, x_n\}$ é ortonormal, resulta daí que

$$\langle z_i, z_j \rangle = \sum_{k=1}^{n} p_{ki} p_{kj} = \delta_{ij}\,,$$

logo $\mathcal{Z}$ é ortonormal e $(7) \Rightarrow (8)$.
Finalmente, se vale (8), seja $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ uma base ortonormal. Para todo $u = \alpha_1 u_1 + \cdots + \alpha_n u_n \in E$ tem-se

$$|u|^2 = \sum_{i=1}^{n} \alpha_i^2.$$

Como $\{Au_1, \ldots, Au_n\} \subset F$ é um conjunto ortonormal,

$$|Au|^2 = \left| \sum_{i=1}^{n} \alpha_i\, Au_i \right|^2 = \sum_{i=1}^{n} \alpha_i^2 = |u|^2,$$

logo $|Au| = |u|$ e $(8) \Rightarrow (1)$. $\square$

Observação. O Teorema 14.1 confirma o dito: enunciado longo $\Rightarrow$ demonstração fácil. (Dito válido com a mesma precisão e generalidade com que costumam valer os provérbios.)

Uma transformação linear $A\colon E \to F$ chama-se *ortogonal* quando cumpre uma das oito condições do Teorema 14.1 (e portanto todas elas). Em particular, um operador linear $A\colon E \to E$ chama-se ortogonal quando $A^* = A^{-1}$. Para que o operador linear $A\colon E \to E$ seja ortogonal, é suficiente que $A^*A = I_E$ ou então que $AA^* = I_E$.

De $|Av| = |v|$ resulta que os únicos autovalores possíveis para um operador ortogonal $A\colon E \to E$ são $+1$ e -1. Com efeito, $Av = \lambda v$ com $v \neq 0$ implica $|v| = |Av| = |\lambda v| = |\lambda|\,|v|$, logo $|\lambda| = 1$.

Se u e v são autovetores do operador ortogonal A, com $Au = u$ e $Av = -v$ então $\langle u, v\rangle = 0$. Com efeito:

$$\langle u, v\rangle = \langle Au, -Av\rangle = \langle A^*Au, -v\rangle = \langle u, -v\rangle = -\langle u, v\rangle\,.$$

Teorema 14.2. *Se o operador ortogonal $A\colon E \to E$ deixa invariante o subespaço $F \subset E$ então A deixa invariante o complemento ortogonal $F^\perp$.*

Demonstração: Dado arbitrariamente $w \in F^\perp$, queremos provar que $Aw \in F^\perp$, isto é, que $\langle Aw, v\rangle = 0$ para todo $v \in F$. Ora, a restrição de A ao subespaço invariante F é um operador injetivo $A\colon F \to F$, logo sobrejetivo. Assim, dado $v \in F$, existe $u \in F$ tal que $v = Au$. Logo $\langle Aw, v\rangle = \langle Aw, Au\rangle = \langle w, u\rangle = 0$, pois $w \in F^\perp$ e $u \in F$. $\square$

Observação. Parte do argumento acima mostra que se o operador $A\colon E \to E$ é invertıvel e $F \subset E$ é invariante por A, então F é invariante por A^{-1}.

Exemplo 14.4. Uma operador linear $S\colon E \to E$ que é ao mesmo tempo ortogonal e auto-adjunto cumpre $S^*S = I$ e $S^* = S$, logo $SS = I$. Portanto ortogonal + auto-adjunto $\Rightarrow$ involução. Pelo Teorema 7.3, tem-se $E = F_1 \oplus F_2$, onde $F_1 = \{v \in E; Sv = v\}$ e $F_2 = \{v \in E; Sv = -v\}$. Assim, os elementos não-nulos de F_1 e F_2 são os autovetores de S correspondentes aos autovalores $+1$ e -1 respectivamente. Da observação que precede o Teorema 14.2 resulta que $F_2 = (F_1)^\perp$. Juntando-se uma base ortonormal de F_1 com uma base ortonormal

de F_2 obtém-se uma base ortonormal de E em relação à qual a matriz de S é diagonal, com a diagonal da forma $(1, \ldots, 1, -1, \ldots, -1)$. O operador S é uma reflexão ortogonal. O leitor pode verificar sem dificuldade que duas quaisquer das seguintes propriedades de um operador linear $S \colon E \to E$ implicam a terceira: (a) $SS = I$; (b) $S^* = S$; (c) $S^* = S^{-1}$.

Em seguida, examinaremos como pode ser um operador ortogonal $A \colon E \to E$ num espaço vetorial *de dimensão* 2, dotado de um produto interno.

Segundo a natureza dos autovalores de A, há quatro possibilidades:

(1) A *possui um único autovalor, o qual é igual a* 1. Neste caso, $A = I$.

Com efeito, seja $u \in E$ um vetor unitário tal que $Au = u$. Se $v \in E$ é outro vetor unitário, perpendicular a u, então Av também é um vetor unitário perpendicular a u, pelo Teorema 14.2, logo $Av = \pm v$. Como A não admite o autovalor -1, deve ser $Av = v$. Mas $\{u, v\} \subset E$ é uma base, portanto $A = I$.

(2) A *possui um único autovalor, o qual é igual a* -1. Então $A = -I$.

O raciocínio é inteiramente análogo ao anterior.

(3) A *admite os autovalores* 1 *e* -1. Então tomamos vetores unitários $u, v \in E$ com $Au = u$ e $Av = -v$, logo $\{u, v\} \subset E$ é uma base ortonormal, relativamente à qual a matriz de A é

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} .$$

Neste caso, A é a reflexão ortogonal em torno do eixo que contém o vetor u.

(4) A *não possui autovalores* (reais).

Então tomamos uma base ortonormal arbitrária $\{u, v\} \subset E$. A matriz de A nesta base, sendo ortogonal 2×2 sem autovalores, tem, segundo o Exemplo 14.2, a forma

$$\mathbf{a} = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} .$$

Somos tentados a dizer que, neste caso, o operador $A \colon E \to E$ é a rotação de ângulo θ. Mas para isso é necessário verificar se θ não

depende da base ortonormal $\mathcal{U} = \{u, v\} \subset E$ escolhida. Ora, se tomarmos outra base ortonormal $\mathcal{U}' = \{u', v'\} \subset E$, a nova matriz de A será $\mathbf{a}' = \mathbf{p}^{-1}\mathbf{a}\mathbf{p}$, onde

$$\mathbf{p} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

matriz de passagem de $\mathcal{U}$ para $\mathcal{U}'$, é ortogonal, logo $\mathbf{p}^{-1} = \mathbf{p}^T$. Assim,

$$\mathbf{a}' = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} .$$

Levando em conta a ortonormalidade das linhas e das colunas da matriz $\mathbf{p}$, um cálculo fácil mostra que se tem

$$\mathbf{a}' = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

ou

$$\mathbf{a}' = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\operatorname{sen}(-\theta) \\ \operatorname{sen}(-\theta) & \cos(-\theta) \end{bmatrix}$$

conforme seja $ad - bc = 1$ ou $ad - bc = -1$ respectivamente.

Noutras palavras, o ângulo θ fica determinado a menos do sinal. Isto quer dizer que, num espaço vetorial de dimensão 2, munido de produto interno, faz sentido a noção de ângulo apenas em valor absoluto. Para que se possa falar no ângulo dotado de sinal é preciso introduzir uma *orientação* em E, isto é, escolher uma base ortonormal $\{u, v\} \subset E$, chamá-la de *positiva* e declarar que são também positivas todas as bases ortonormais de E que se obtêm a partir desta por meio de matrizes de passagem ortogonais cujo determinante $ad - bc$ é igual a 1. De qualquer modo, com ou sem orientação, os operadores ortogonais sem autovalores (juntamente com $\pm I$) num espaço vetorial de dimensão 2 serão chamados *rotações*. (I e $-I$ são respectivamente as rotações de ângulo $0°$ e $180°$ respectivamente.)

Teorema 14.3. *Seja $A: E \to E$ um operador ortogonal num espaço vetorial de dimensão finita munido de produto interno. Existe uma*

base ortonormal de E *relativamente à qual a matriz de* A *tem a forma*

$$\begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & -1 & & & & & & & \\ & & & & \ddots & & & & & & \\ & & & & & -1 & & & & & \\ & & & & & & \cos\alpha_1 & -\operatorname{sen}\alpha_1 & & & \\ & & & & & & \operatorname{sen}\alpha_1 & \cos\alpha_1 & & & \\ & & & & & & & & \ddots & & \\ & & & & & & & & & \cos\alpha_k & -\operatorname{sen}\alpha_k \\ & & & & & & & & & \operatorname{sen}\alpha_k & \cos\alpha_k \end{bmatrix},$$

onde os termos não aludidos são iguais a zero.

Demonstração: Os conjuntos $F_1 = \{v \in E; Av = v\}$ e $F_2 = \{v \in E; Av = -v\}$ são subespaços vetoriais invariantes por A, com $F_1 \cap F_2 = \{0\}$. Logo $F = F_1 \oplus F_2$ também é um subespaço invariante por A, o mesmo acontecendo com o complemento ortogonal $F^\perp$. Seja $\{u_1, \ldots, u_r\} \subset F$ uma base ortonormal cujos primeiros vetores formam uma base de F_1 e os restantes uma base de F_2. Nenhum subespaço de dimensão 1 em $F^\perp$ é invariante por A, logo existe um subespaço invariante $G \subset F^\perp$ de dimensão 2. Seja $\{v_1, w_1\} \subset G$ uma base ortonormal. Como vimos acima, tem-se $Av_1 = \cos\alpha_1 \cdot v_1 + \operatorname{sen}\alpha_1 \cdot w_1$, $Aw_1 = -\operatorname{sen}\alpha_1 \cdot v_1 + \cos\alpha_1 \cdot w_1$. Novamente, o complemento ortogonal de G em $F^\perp$ é um subespaço invariante por A, que não possui autovetores, logo contém um subespaço invariante de dimensão 2. Prosseguindo analogamente, chegaremos a uma base ortonormal $\{v_1, w_1, \ldots, v_k, w_k\} \subset F^\perp$ tal que $Av_i = \cos\alpha_i v_i + \operatorname{sen}\alpha_i v_i$ e $Aw_i = -\operatorname{sen}\alpha_i v_i + \cos\alpha_i w_i$ para $i = 1, \ldots, k$. Então $\{u_1, \ldots, u_r, v_1, w_1, \ldots, v_k, w_k\} \subset E$ é uma base ortonormal, relativamente à qual a matriz de A tem a forma desejada. □

Observação. A matriz a que se refere o Teorema 14.3 pode não possuir elementos iguais a 1, ou a -1, como pode também não conter nenhum dos blocos 2×2, característicos de rotações. Além disso, caso seja conveniente, cada bloco igual a $\mathbf{I}_2$ ou a $-\mathbf{I}_2$ pode ser substituído

respectivamente pelos blocos

$$\begin{bmatrix} \cos 0 & -\operatorname{sen} 0 \\ \operatorname{sen} 0 & \cos 0 \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \cos \pi & -\operatorname{sen} \pi \\ \operatorname{sen} \pi & \cos \pi \end{bmatrix}.$$

Corolário. *Se E tem dimensão ímpar, todo operador ortogonal $A: E \to E$ possui um autovetor v, com $Av = v$ ou $Av = -v$.*

Exemplo 14.5. Seja $A: \mathbb{R}^3 \to \mathbb{R}^3$ um operador ortogonal no espaço euclidiano tri-dimensional. Se $A \neq \pm I$ então existe uma base ortonormal $\{u_1, u_2, u_3\} \subset \mathbb{R}^3$ em relação à qual a matriz de A tem uma das quatro formas abaixo:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\operatorname{sen} \alpha \\ 0 & \operatorname{sen} \alpha & \cos \alpha \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 & 0 \\ 0 & \cos \alpha & -\operatorname{sen} \alpha \\ 0 & \operatorname{sen} \alpha & \cos \alpha \end{bmatrix}.$$

No primeiro caso A é a reflexão em torno do plano que contém os vetores u_1, u_2 (paralelamente a u_3). No segundo caso, A é a reflexão em torno do eixo que contém u_1 ou, o que é o mesmo, a rotação de $180°$ em torno desse eixo. No terceiro caso, A é a rotação de ângulo α em torno do eixo que contém u_1. No último caso, A é essa rotação seguida da reflexão em torno do plano que contém u_2 e u_3. Estas quatro possibilidades, mais $A = I$ e $A = -I$, esgotam todos os tipos de operadores ortogonais em $\mathbb{R}^3$.

Teorema 14.4. *Seja E um espaço vetorial de dimensão finita, munido de produto interno. Todo operador linear $A: E \to E$ admite uma decomposição da forma $A = PU$, onde $U: E \to E$ é ortogonal e $P: E \to E$ é não-negativo.*

Demonstração: De acordo com o Teorema 13.10, existem bases ortonormais $\{u_1, \ldots, u_n\} \subset E$, $\{v_1, \ldots, v_n\} \subset E$ e números $\lambda_i \geq 0$ tais que $Au_i = \lambda_i v_i$ para $i = 1, \ldots, n$. Definamos os operadores $P, U: E \to E$ impondo que $Pv_i = \lambda_i v_i$ e $Uu_i = v_i$. Evidentemente, P é auto-adjunto, ≥ 0, U é ortogonal e $PU = A$. □

A expressão $A = PU$ chama-se uma *decomposição polar* do operador A, por analogia com a forma polar $z = re^{i\theta}$ de um número

complexo. Os fatores P e U são univocamente determinados a partir de A, no caso em que A é invertível. Provemos isto. É claro que A invertível obriga P a ser também invertível, donde positivo. De $A = PU$ tiramos $A^* = U^*P$ e, multiplicando membro a membro estas igualdades, vem $AA^* = PUU^*P = P^2$, portanto P é a raiz quadrada positiva do operador positivo AA^*. (Vide Teorema 13.8.) Multiplicando a igualdade $A = PU$ por P^{-1} à esquerda, vem $U = P^{-1}A$. Isto mostra que, no caso em que A é invertível, $P = \sqrt{AA^*}$ e $U = P^{-1}A$ são univocamente determinados a partir da igualdade $A = PU$.

Exercícios

14.1. Prove que as linhas de uma matriz $\mathbf{a}$ são duas a duas ortogonais se, e somente se, $\mathbf{a}\mathbf{a}^T = \mathbf{d}$, onde $\mathbf{d}$ é uma matriz diagonal. Enuncie e prove um resultado análogo sobre a ortogonalidade das colunas de $\mathbf{a}$.

14.2. Se as linhas de uma matriz quadrada forem duas a duas ortogonais e tiverem a mesma norma, prove que as colunas dessa matriz também são duas a duas ortogonais.

14.3. Dê os seguintes exemplos:

(a) Uma matriz invertível cujas linhas são duas a duas ortogonais mas as colunas não são.

(b) Uma matriz (não-quadrada) cujas linhas são ortogonais e têm a mesma norma mas as colunas não são ortogonais.

(c) Uma matriz cujas linhas (e colunas) são duas a duas ortogonais mas as normas das linhas são diferentes.

14.4. Seja $f \colon \mathbb{R}^n \to \mathbb{R}^n$ uma função tal que $f(0) = 0$ e $|f(u) - f(v)| = |u - v|$ para quaisquer $u, v \in \mathbb{R}^n$. (Ou seja: f deixa 0 fixo e preserva distâncias.) Prove:

(a) Para todo $v \in \mathbb{R}^n$, tem-se $|f(v)| = |v|$.

(b) Para quaisquer $u, v \in \mathbb{R}^n$, tem-se $\langle f(u), f(v) \rangle = \langle u, v \rangle$. [Use a igualdade $\langle u, v \rangle = \frac{1}{2}(|u|^2 + |v|^2 - |u - v|^2)$.]

(c) Os vetores $u_1 = f(e_1), \ldots, u_n = f(e_n)$ formam uma base ortonormal em $\mathbb{R}^n$.

(d) Para todo $v = x_1e_1 + \cdots + x_ne_n \in \mathbb{R}^n$, tem-se $\langle f(v), u_i\rangle = x_i$, logo $f(v) = x_1u_1 + \cdots + x_nu_n$.

(e) $f\colon \mathbb{R}^n \to \mathbb{R}^n$ é um operador linear, logo é ortogonal.

Uma função $g\colon \mathbb{R}^n \to \mathbb{R}^n$ chama-se uma *isometria* quando $|g(u) - g(v)| = |u - v|$ para quaisquer $u, v \in \mathbb{R}^n$. Conclua que toda isometria tem a forma $g(v) = A \cdot v + b$, onde $A\colon \mathbb{R}^n \to \mathbb{R}^n$ é um operador (linear) ortogonal e $b \in \mathbb{R}^n$ é um vetor constante (independente de v).

14.5. Seja E um espaço vetorial de dimensão finita, com produto interno. Se $A_o\colon F \to E$ é uma transformação linear ortogonal definida num subespaço vetorial $F \subset E$, prove que existe um operador ortogonal $A\colon E \to E$ tal que $A \cdot v = A_o \cdot v$ para todo $v \in F$.

14.6. Sejam $A\colon F \to G$ e $B\colon F \to H$ transformações lineares invertíveis. (G e H são espaços de dimensão finita, munidos de produto interno.) Prove que existe uma transformação ortogonal (invertível) $C\colon G \to H$ com $B = CA$ se, e somente se, $|Av| = |Bv|$ para todo $v \in F$.

14.7. Seja E um espaço vetorial de dimensão finita, com produto interno. Dados dois operadores $A, B\colon E \to E$ tais que $|Av| = |Bv|$ para todo $v \in E$, prove que existe $C\colon E \to E$ ortogonal, com $B = CA$. (Sugestão: observe que $\mathcal{N}(A) = \mathcal{N}(B)$. Considere $F = \mathcal{N}(A)^\perp$ e sejam $A_o\colon F \to \mathcal{I}m(A)$, $B_o\colon F \to \mathcal{I}m(B)$ os isomorfismos obtidos por restrição de A e B respectivamente. Use o exercício anterior para achar $C_o\colon \mathcal{I}m(A) \to \mathcal{I}m(B)$, com $B_o = C_oA_o$ e obtenha C pelo Exercício 14.5.)

14.8. Dada uma base ortonormal $\{u_1, u_2, u_3\} \subset \mathbb{R}^3$, sejam $n, p \in \mathbb{N}$ tais que $p = n^2 + n + 1$. Defina um operador $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ pondo $Au_1 = v_1$, $Au_2 = v_2$, $Au_3 = v_3$, onde

$$v_1 = \frac{1}{p}[nu_1 + (n+1)u_2 + n(n+1)u_3],$$
$$v_2 = \frac{1}{p}[n(n+1)u_1 + nu_2 - (n+1)u_3],$$
$$v_3 = \frac{1}{p}[(n+1)u_1 - n(n+1)u_2 + nu_3].$$

Prove que o operador A é ortogonal.

14.9. Para quaisquer duas bases ortonormais $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ e $\mathcal{V} = \{v_1, \ldots, v_n\} \subset E$, prove que existe um operador ortogonal $A\colon E \to E$ tal que $Au_1 = v_1, \ldots, Au_n = v_n$. Se as bases dadas são formadas pelos vetores

$$u_1 = \frac{1}{3}(1,2,2), \quad u_2 = \frac{1}{3}(2,1,-2), \quad u_3 = \frac{1}{3}(2,-2,1) \quad \text{e}$$
$$v_1 = \frac{1}{7}(2,3,6), \quad v_2 = \frac{1}{7}(6,2,-3), \quad v_3 = \frac{1}{7}(3,-6,2) \text{ em } \mathbb{R}^3,$$

determine a matriz de A na base canônica de $\mathbb{R}^3$.

14.10. Dado o vetor unitário $u \in \mathbb{R}^n$, prove que o operador $H_u\colon \mathbb{R}^n \to \mathbb{R}^n$, definido por $H_u \cdot v = v - 2\langle v, u\rangle\, u$, é ortogonal. (Reflexão em torno de $\{u\}^\perp$.) Dados os vetores $v \neq w$ em $\mathbb{R}^n$, com $|v| = |w|$, mostre que, tomando $u = (v-w)/|v-w|$, tem-se $H_u v = w$. Determine a matriz de H_u em função das coordenadas de u ("matriz de Householder").

14.11. Prove que todo operador $A\colon E \to E$, num espaço vetorial de dimensão finita munido de produto interno se escreve como $A = UP$, onde U é ortogonal e P é não-negativo. (Faça a decomposição polar de A^*.)

14.12. Com a notação do Exercício 11.17, considere um operador ortogonal $A \in \mathcal{L}(E)$ e defina $M_A\colon \mathcal{L}(E) \to \mathcal{L}(E)$ pondo $M_A \cdot X = AX$. Prove que M_A é um operador ortogonal.

14.13. Se uma matriz ortogonal é triangular, prove que ela é diagonal e seu quadrado é igual à matriz identidade.

14.14. Seja E um espaço vetorial de dimensão finita, com produto interno. Uma função $S\colon E \to E$ chama-se uma *semelhança* quando existe um número $r > 0$ (chamado a *razão de semelhança*) tal que $|S(u) - S(v)| = r|u - v|$ para quaisquer $u, v \in E$. Se S é uma semelhança de razão r, prove que existem um operador ortogonal $A\colon E \to E$ e um vetor $b \in E$ tais que $S(v) = r.Av + b$ para todo $v \in E$.

14.15. Seja $A\colon E \to E$ um operador linear que transforma vetores unitários em vetores unitários. Prove que A é ortogonal. Deduza daí que se $S\colon E \to E$ é um operador linear invertível que transforma dois quaisquer vetores de mesmo comprimento em vetores de mesmo comprimento então S é uma semelhança.

14.16. Seja $S\colon E \to E$ um operador linear invertível que preserva ângulos, isto é

$$\frac{\langle Su, Sv\rangle}{|Su|\,|Sv|} = \frac{\langle u, v\rangle}{|u|\,|v|}$$

quando $u \neq 0$ e $v \neq 0$. Prove que S transforma vetores ortogonais de mesmo comprimento em vetores ortogonais de igual comprimento. Conclua que S é uma semelhança.

14.17. Com o produto interno introduzido no Exercício 11.17, prove que os operadores ortogonais têm norma igual a $\sqrt{n}$, onde $n = \dim E$.

14.18. Se a decomposição polar de um operador é única, prove que esse operador é invertível.

14.19. Seja $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ dado por $A(x, y, z) = (2x + 3y - 6z, 6x + 2y + 3z, -3x + 6y + 2z)$. Mostre que A é uma semelhança de razão 7. Sabe-se que ou existe $v \in \mathbb{R}^3$ com $Av = 7v$ ou existe $w \in \mathbb{R}^3$ com $Aw = -7w$. Ache um autovetor de A, complete-o de modo a obter uma base ortonormal de $\mathbb{R}^3$ e determine a matriz do operador A nesta base.

14.20. Seja $\mathbf{a} = [a_1\, a_2\, \ldots\, a_n] \in M(1 \times n)$ tal que $a_1^2 + \cdots + a_n^2 = 1$. Prove que $\mathbf{a}^T\mathbf{a} \in M(n \times n)$ é a matriz de uma projeção ortogonal. Determine a imagem e o núcleo dessa projeção.

14.20. Pode uma matriz ortogonal ser anti-simétrica?

14.21. Seja **a** uma matriz ortogonal $n \times n$.

(a) Prove que $A\colon M(n \times n) \to M(n \times n)$, definida por $A\mathbf{x} = \mathbf{a}\mathbf{x}^T + \mathbf{x}\mathbf{a}^T$, é uma transformação linear cuja imagem é o conjunto das matrizes simétricas.

(b) Prove que, dada uma matriz simétrica $\mathbf{s} \in M(n \times n)$, o conjunto das matrizes $\mathbf{x}$ tais que $\mathbf{a}\mathbf{x}^T + \mathbf{x}\mathbf{a}^T = \mathbf{s}$ é uma variedade afim de dimensão $n(n-1)/2$ no espaço vetorial $M(n \times n)$.

14.22. Ache uma matriz ortogonal 4×4 cujos elementos são todos da forma $\pm\frac{1}{2}$.

14.23. Seja $v = (a, b, c)$ um vetor unitário diferente de $\pm e_3$. Mostre que existe x tal que a matriz abaixo é ortogonal:

$$\begin{bmatrix} a & b & c \\ bx & -ax & 0 \\ acx & bcx & -1/x \end{bmatrix}.$$

14.24. Um operador auto-adjunto $N \colon E \to E$ chama-se *negativo* quando $\langle Nv, v \rangle < 0$ para todo $v \neq 0$ em E. Determine a (única) decomposição polar de um operador negativo N.

14.25. Prove que os elementos da diagonal de uma matriz negativa são números negativos.

14.26. Prove que a matriz abaixo é negativa $\begin{bmatrix} -34 & 12 \\ 12 & -41 \end{bmatrix}$.

14.27. Ache a decomposição polar da matriz $\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$.

14.28. Sejam $\mathbf{a} \in M(r \times r)$ e $\mathbf{c} \in M(s \times s)$ matrizes ortogonais. Prove que a matriz $(r + s) \times (r + s)$ abaixo é ortogonal se, e somente se, $\mathbf{b} = \mathbf{0}$:

$$\begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} \end{bmatrix}.$$

14.29. Obtenha a decomposição polar da matriz

$$\begin{bmatrix} \sqrt{2} & 1 & 1 \\ -\sqrt{2} & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

15

Operadores Normais (Caso Real)

Estudaremos agora o tipo mais geral de operadores (num espaço vetorial de dimensão finita, munido de produto interno) para os quais vale um resultado análogo ao do Teorema 14.3, ou seja: existe uma base ortonormal na qual a matriz do operador começa na forma diagonal, seguida por uma série de blocos 2×2 da forma

$$\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

Eles são os operadores normais. Na Seção 21 veremos que, nos espaços vetoriais complexos, os operadores normais são precisamente aqueles que admitem uma base ortonormal de autovetores.

Nesta seção, todos os espaços vetoriais têm dimensão finita e são munidos de produto interno.

Um operador linear $A\colon E \to E$ chama-se *normal* quando comuta com seu adjunto, isto é, quando $AA^* = A^*A$.

Uma matriz quadrada $\mathbf{a}$ diz-se *normal* quando comuta com sua transposta, isto é, quando $\mathbf{a}\mathbf{a}^T = \mathbf{a}^T\mathbf{a}$. Portanto um operador é normal se, e somente se, sua matriz relativamente a uma base ortonormal é uma matriz normal.

Exemplo 15.1. Os operadores auto-adjuntos e os ortogonais são normais. Analogamente, as matrizes simétricas e as ortogonais quadradas são normais.

Exemplo 15.2. (Matrizes normais 2×2.) Seja $\mathbf{a} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ uma matriz normal 2×2. Temos

$$\mathbf{a}\mathbf{a}^T = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}$$

e

$$\mathbf{a}^T\mathbf{a} = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix}.$$

Logo $\mathbf{a}$ é normal se, e somente se, $b^2 = c^2$ (isto é, $b = \pm c$) e $ab + cd = ac + bd$. Se $b = c$, a matriz $\mathbf{a}$ é simétrica. Caso contrário, temos $c = -b$ (com $b \neq 0$). Então, de $ab + cd = ac + bd$ resulta $b(a - d) = b(d - a)$, donde $a = d$. Portanto, as únicas matrizes normais 2×2 são as simétricas e as da forma

$$\mathbf{a} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

Observe que se $\mathbf{a} \neq 0$ então $0 \neq r = \sqrt{a^2 + b^2}$. Logo existe $\theta \in \mathbb{R}$ tal que $\cos\theta = a/r$ e $\operatorname{sen}\theta = b/r$. Então a matriz $\mathbf{a}$ se escreve como

$$\mathbf{a} = r \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Isto nos permite concluir que uma matriz normal 2×2 ou é simétrica ou é a matriz de uma semelhança no plano.

Um operador linear $A \colon E \to E$ chama-se *anti-simétrico* quando $A^* = -A$, ou seja, $\langle Au, v\rangle = -\langle u, Av\rangle$ para $u, v \in E$ quaisquer. Para que A seja anti-simétrico é necessário e suficiente que sua matriz $[a_{ij}]$ em relação a uma base ortornormal de E seja anti-simétrica, isto é, $a_{ij} = -a_{ji}$ para quaisquer $i, j = 1, 2, \ldots, n$. Em particular, $a_{ii} = 0$. Num espaço de dimensão 1, todo operador anti-simétrico é igual a zero; logo o único autovalor possível de um operador anti-simétrico é 0.

Evidentemente, todo operador anti-simétrico é normal.

Teorema 15.1. *Seja* $A\colon E \to E$ *um operador normal. Existe uma base ortonormal de* E *na qual a matriz de* A *tem a forma*

$$\begin{bmatrix} \lambda_1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda_r & & & & & & \\ & & & \alpha_1 & -\beta_1 & & & & \\ & & & \beta_1 & \alpha_1 & & & & \\ & & & & & \ddots & & & \\ & & & & & & \alpha_s & -\beta_s \\ & & & & & & \beta_s & \alpha_s \end{bmatrix}$$

onde os elementos não expostos são iguais a zero.

Demonstração: Sejam $\sigma_0^2, \sigma_1^2, \dots, \sigma_k^2$ os autovalores distintos do operador não-negativo $AA^* = A^*A\colon E \to E$, com $\sigma_0 = 0$ e $\sigma_i > 0$ se $i \geq 1$. Para cada $i = 0, 1, \dots, k$, seja $E_i = \mathcal{N}(AA^* - \sigma_i^2 I)$ o auto-subespaço correspondente ao autovalor σ_i^2. O Teorema Espectral nos afirma que $E = E_0 \oplus E_1 \oplus \cdots \oplus E_k$. Além disso, se $u \in E_i$ e $v \in E_j$ com $i \neq j$ então $\langle u, v\rangle = 0$. Da igualdade $AA^* = A^*A$ resulta imediatamente que cada subespaço E_i é invariante por A e por A^*. Ora, por definição, AA^* em E_i coincide com $\sigma_i^2 I$. Logo o operador $B_i\colon E_i \to E_i$, definido para $i = 1, 2, \dots, k$ por $B_i v = (1/\sigma_i)Av$, é ortogonal. Tomemos em E_0 uma base ortonormal qualquer. (Observe que $E_0 = \mathcal{N}(AA^*) = \mathcal{N}(A) = \mathcal{N}(A^*)$.) Tomemos ainda, em cada subespaço E_i $(i = 1, \dots, k)$ uma base ortonormal na qual a matriz do operador B_i tenha a forma dada no Teorema 14.3. Juntando todas essas bases e ordenando seus elementos na forma adequada, obtemos uma base de E na qual a matriz do operador A é do tipo desejado.

Observações:

1. Os números $\lambda_1, \dots, \lambda_r$ são os autovalores de A. Dentre eles, os não-nulos têm a forma $\lambda_i = \pm\sigma_j$, onde σ_j é um valor singular de A. Os demais valores singulares de A são $\sigma_j = \sqrt{\alpha_j^2 + \beta_j^2}$ $(j = 1, \dots, s)$.

2. Se A não possui autovetores, a matriz do Teorema 15.1 não apresenta a parte superior (diagonal). Por outro lado, se A é auto-adjunto não existem os blocos 2×2 da parte inferior.

Corolário. *Seja* $A\colon E \to E$ *um operador anti-simétrico. Existe uma base ortonormal de* E *relativamente à qual a matriz de* A *tem a forma*

$$\begin{bmatrix} 0 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 0 & & & & & & \\ & & & 0 & -\beta_1 & & & & \\ & & & \beta_1 & 0 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 0 & -\beta_s \\ & & & & & & \beta_s & 0 \end{bmatrix}$$

Segue-se imediatamente que o posto de um operador anti-simétrico é um número par.

Exemplo 15.4. Seja $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ um operador anti-simétrico. Pelo corolário acima, existe uma base ortonormal $\{u_1, u_2, u_3\} \subset \mathbb{R}^3$ tal que $Au_1 = 0$, $Au_2 = -\beta u_3$ e $Au_3 = \beta u_2$. Em termos do produto vetorial clássico de $\mathbb{R}^3$, podemos escolher o sinal de u_1 de modo que se tenha $u_1 \times u_2 = u_3$ e $u_3 \times u_1 = u_2$. Então, se pusermos $w = \beta u_1$, teremos $0 = Au_1 = u_1 \times w$, $Au_2 = -\beta u_3 = \beta(u_2 \times u_1) = u_2 \times w$ e $Au_3 = \beta u_2 = \beta(u_3 \times u_1) = u_3 \times w$. Assim, A coincide com o operador linear $v \mapsto v \times w$ nos vetores básicos u_1, u_2, u_3. Concluímos que, para todo $v \in \mathbb{R}^3$, vale $Av = v \times w$. Em resumo: todo operador anti-simétrico no espaço $\mathbb{R}^3$ consiste no produto vetorial por um vetor fixo $w \in \mathbb{R}^3$.

Exercícios

15.1. Seja $A = B + C$ a decomposição do operador A como soma do operador auto-adjunto B com o operador anti-simétrico C. Prove que A é normal se, e somente se, $BC = CB$.

15.2. Prove que os operadores auto-adjuntos $S, T\colon E \to E$ são iguais se, e somente se, $\langle Sv, v\rangle = \langle Tv, v\rangle$ para todo $v \in E$. Use este fato para provar que $A\colon E \to E$ é normal se, e somente se, $|Av| = |A^*v|$ para todo $v \in E$.

15.3. Se $\dim E = n$ e o operador normal $A\colon E \to E$ tem n autovalores distintos, prove que A é auto-adjunto.

15.4. Sejam $u_1, \ldots, u_n$ as linhas e $v_1, \ldots, v_n$ as colunas de uma matriz **a**. Prove que $\langle u_i, u_j \rangle = \langle v_i, v_j \rangle$ para quaisquer $i, j \Leftrightarrow$ **a** é normal.

15.5. Prove que uma projeção $P\colon E \to E$ é um operador normal se, e somente se, $P = P^*$. Resultado análogo para uma involução.

15.6. Seja $A\colon E \to E$ um operador normal. Prove que $\mathcal{N}(A) = \mathcal{N}(A^*)$ e conclua que $\mathcal{N}(A)^\perp = \mathcal{I}m(A) = \mathcal{I}m(A^*)$.

15.7. Entre as matrizes abaixo, determine quais são normais.

$$\mathbf{a} = \begin{bmatrix} 9 & -3 & -6 \\ 3 & 9 & 6 \\ 6 & -6 & 9 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \\ 2 & 3 & 5 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 2 \\ 0 & -2 & -1 \end{bmatrix}$$

15.8. Sejam $A, B\colon E \to E$ operadores normais. Supondo $AB = 0$, prove:

(a) A imagem de B está contida no núcleo de A;

(b) A imagem de B^* está contida no núcleo de A^*;

(c) A imagem de A está contida no núcleo de B.

Conclua então que $BA = 0$.

15.9. Se $A, B\colon E \to E$ são normais e $AB = BA$, prove que AB é normal.

15.10. Dê exemplos de operadores normais A, B tais que $A + B$ não é normal e AB não é normal. Dê também um exemplo em que AB é normal mas $AB \neq BA$.

15.11. Seja $A\colon E \to E$ um operador linear no espaço E, de dimensão finita, com produto interno. Supondo $I + A$ e $I - A$ invertíveis, defina $S = (I + A)(I - A)^{-1}$. Prove que S é ortogonal se, e somente se, A é anti-simétrico.

15.12. Seja o operador $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ dado por $A(x, y, z) = (y + 2z, -x + 3z, -2x - 3y)$. Escreva sua matriz na base canônica e veja que A é anti-simétrico. Por escalonamento, ache um vetor $u \in \mathbb{R}^3$ que é uma base de $\mathcal{N}(A)$. Obtenha ainda uma base $\{v, w\} \subset \mathcal{N}(A)^\perp$ e escreva a matriz de A na base $\{u, v, w\} \subset \mathbb{R}^3$.

15.13. Seja $A\colon \mathbb{R}^4 \to \mathbb{R}^4$ o operador linear definido por

$$A(x,y,z,t) = (y - z + t, -x - z + 2t, x + y - t, -x - 2y + z).$$

Mostre que A é anti-simétrico. Ache bases ortonormais $\{u, v\} \subset \mathcal{N}(A)$ e $\{u', v'\} \subset \mathcal{N}(A)^{\perp}$. Determine a matriz de A na base $\{u, v, u', v'\} \subset \mathbb{R}^4$.

15.14. Se o operador $A\colon E \to E$ é anti-simétrico, prove que todo vetor $v \in E$ é perpendicular à sua imagem Av. Se $E = \mathbb{R}^2$, prove que A é um múltiplo αR da rotação de $90°$ $R\colon \mathbb{R}^2 \to \mathbb{R}^2$.

15.15. Seja

$$\mathbf{a} = \begin{bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{bmatrix}$$

a matriz do operador $A\colon \mathbb{R}^3 \to \mathbb{R}^3$. Mostre que, para todo $v \in \mathbb{R}^3$ tem-se $Av = v \times w$, onde $w = (a, b, c)$.

15.16. Seja $A\colon E \to E$ um operador normal. Se $Au = \lambda u$ e $Av = \mu v$ com $\lambda \neq \mu$, prove que $\langle u, v \rangle = 0$.

15.17. Sejam $u_1, ..., u_n$ os vetores-linha e $v_1, ..., v_n$ os vetores-coluna da matriz $\begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} \end{bmatrix}$, onde $\mathbf{a} \in M(p \times p)$. Se $|u_i| = |v_i|$ para $i = 1, \ldots, n$, prove que $\mathbf{b} = \mathbf{0} \in M(p \times (n - p))$.

15.18. Se um subespaço $F \subset E$ é invariante pelo operador normal $A\colon E \to E$, prove que F também é invariante por A^* e que o complemento ortogonal $F^{\perp}$ é ainda invariante por A e A^*.

15.19. Seja $F \subset E$ um subespaço invariante pelo operador normal $A\colon E \to E$. Prove que a restrição de A ao subespaço F é ainda um operador normal.

16

Pseudo-inversa

A noção de pseudo-inversa e o estudo de suas propriedades constituem uma maneira simples e atraente de aplicar alguns dos resultados obtidos nas seções anteriores. Do ponto de vista prático, esta noção responde uma pergunta bastante natural, que ocorre de fato em diferentes aplicações da Álgebra Linear: dada $A \in \mathcal{L}(E; F)$ *e dado* $b \in F$*, se é impossível achar* $x \in E$ *tal que* $Ax = b$*, qual é, ou melhor, quais são os vetores* $x \in E$ *tais que o erro* $|Ax - b|$ *é o menor possível e qual entre esses vetores* x *é a solução ótima, ou seja, tem a menor norma?*

Sabemos que um sistema de m equações lineares com n incógnitas pode ser interpretado como o problema de achar um vetor $x \in \mathbb{R}^n$ tal que $Ax = b$, onde $A\colon \mathbb{R}^n \to \mathbb{R}^m$ é a transformação linear cuja matriz (nas bases canônicas de $\mathbb{R}^n$ e $\mathbb{R}^m$) é dada pelos coeficientes do sistema e $b \in \mathbb{R}^m$ é o vetor cujas coordenadas são os números que figuram nos segundos membros das equações do sistema.

Se b não pertence à imagem de A, o sistema $Ax = b$ evidentemente não possui solução. Faz sentido, entretanto, procurar em $\mathbb{R}^n$ um vetor x tal que Ax esteja o mais próximo possível de b e, dentre esses vetores x, aquele de menor norma. Isto nos leva à noção de pseudo-inversa de uma transformação linear.

Seja $A\colon E \to F$ uma transformação linear entre espaços vetoriais de dimensão finita, munidos de produto interno. A *pseudo-inversa*

de A é a correspondência $A^+\colon F \to E$ que associa a cada $y \in F$ o vetor $A^+y = x \in E$ de menor norma entre todos os vetores $x \in E$ que tornam mínima a distância $|y - Ax|$. (Fig. 16.1.)

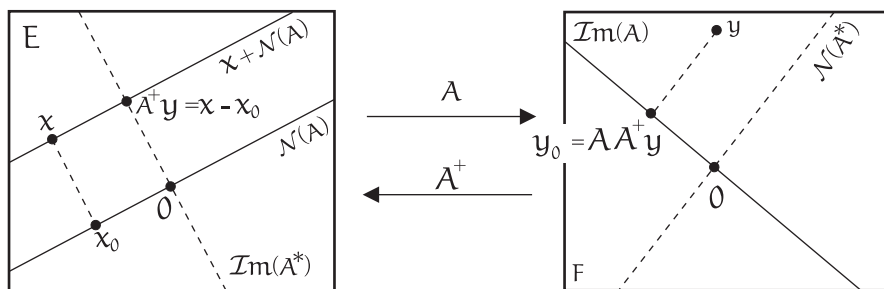



Figura 16.1.

Descrevamos como opera a pseudo-inversa $A^+\colon F \to E$. Dado $y \in F$, o vetor de $\mathcal{I}m(A)$ mais próximo de y é a projeção ortogonal y_o de y sobre $\mathcal{I}m(A)$, caracterizada pelo fato de que $y_o \in \mathcal{I}m(A)$ e $y - y_o$ é perpendicular a (todos os vetores de) $\mathcal{I}m(A)$, ou seja, $y - y_o \in \mathcal{N}(A^*)$. Como $y_o \in \mathcal{I}m(A)$, existem vetores $x \in E$ tais que $Ax = y_o$. Se x é um deles, os demais são da forma $x + z$, onde $z \in \mathcal{N}(A)$, pelo Teorema 6.4. Dentre estes vetores $x + z$, o de menor norma é $x - x_o$, onde x_o é a projeção ortogonal de x sobre $\mathcal{N}(A)$ pois sendo $x - x_o$ perpendicular a $\mathcal{N}(A)$, Pitágoras nos dá, para todo $z \in \mathcal{N}(A)$:

$$|x+z|^2 = |x - x_o + z + x_o|^2 = |x - x_o|^2 + |z + x_o|^2 \geq |x - x_o|^2.$$

(pois $z + x_o \in \mathcal{N}(A)$, logo é perpendicular a $x - x_o$). Portanto $A^+y = x - x_o$. Note que, sendo ortogonal a $\mathcal{N}(A)$, A^+y pertence a $\mathcal{I}m(A^*)$. Na realidade, A^+y é o único vetor da imagem de A^* tal que $AA^+y = y_o$. (Com efeito, A, restrita a $\mathcal{I}m(A^*)$, é injetiva, visto que $\mathcal{I}m(A^*) \cap \mathcal{N}(A) = \{0\}$.)

Esta definição da pseudo-inversa de uma transformação linear $A\colon E \to F$ apresenta-a como uma função bem definida $A^+\colon F \to E$, com as propriedades geométricas procuradas, porém não deixa claro se A^+ é uma transformação linear. Usando o Teorema 13.10, apresentaremos em seguida uma transformação $A'\colon F \to E$, que certamente é linear mas que usa certas bases escolhidas em E e F, de

modo que não parece estar bem definida. Em seguida, provaremos que $A' = A^+$, logo A^+ é linear e A' não depende das escolhas de bases.

Seja r o posto da transformação linear $A: E \to F$. Pelo Teorema 13.10, existem bases ortonormais $\{u_1, \ldots, u_n\} \subset E$, $\{v_1, \ldots, v_m\} \subset F$ e números positivos $\sigma_1, \ldots, \sigma_r$ tais que $Au_i = \sigma_i v_i$, $A^* v_i = \sigma_i u_i$ para $1 \le i \le r$ e $Au_i = 0$, $A^* v_i = 0$ para $i > r$. Pelo Teorema 4.1, existe uma única transformação linear $A': F \to E$ tal que

$$A' v_i = \frac{1}{\sigma_i} u_i \text{ para } 1 \le i \le r \quad \text{e} \quad A' v_i = 0 \text{ quando } i > r.$$

Teorema 16.1. $A' = A^+ =$ *pseudo-inversa de* A.

Demonstração: Devemos mostrar que, para todo $y \in F$, o vetor $A'y$ é igual a $A^+ y$, isto é, tem as seguintes propriedades: (1) $AA'y = y_o$ é o vetor de $\mathcal{I}m(A)$ mais próximo de y; (2) $A'y$ é o vetor de menor norma em E cuja imagem por A é y_o. Estas afirmações são equivalentes a $(1')$ $(y - AA'y) \perp \mathcal{I}m(A)$, ou seja, $y - AA'y \in \mathcal{N}(A^*)$, ou ainda, $A^* y = A^* AA'y$; $(2')$ $A'y \in \mathcal{I}m(A^*)$. Evidentemente, basta verificar a validez de $(1')$ e $(2')$ quando y é qualquer um dos vetores básicos $v_1, \ldots, v_m$. Nestes casos, porém, $(1')$ e $(2')$ resultam imediatamente da definição de A'. □

Corolário 1. $AA^+: F \to F$ *é a projeção ortogonal sobre* $\mathcal{I}m(A)$ *e* $A^+ A: E \to E$ *é a projeção ortogonal sobre* $\mathcal{I}m(A^*)$.

Com efeito, temos $AA'v_i = v_i$ se $1 \le i \le r$, $A'v_i = 0$ se $i > r$, $A' = A^+$ e $\{v_1, \ldots, v_r\} \subset \mathcal{I}m(A)$ é uma base. Analogamente para A^+A. (O Corolário 1 também resulta diretamente da definição de A^+.) □

Corolário 2. *Se* $A: E \to F$ *é injetiva então* $A^+ = (A^*A)^{-1} A^*$.

Com efeito, se A é injetiva, o operador $A^*A: E \to E$ é invertível. (Corolário do Teorema 13.9). A igualdade alegada significa que $A^*(AA^+) = A^*$, o que foi estabelecido na prova do Teorema 16.1. (Forma final da condição $(1')$.) □

Corolário 3. *Se* $A: E \to F$ *é sobrejetiva então* $A^+ = A^*(AA^*)^{-1}$.

Com efeito, ainda pelo Corolário do Teorema 13.9, A sobrejetiva implica que o operador $AA^*: F \to F$ é invertível e a igualdade alegada

equivale a $(A^+A)A^* = A^*$. Isto é evidente se substituirmos A^+ por A' (vide prova do Teorema 16.1) e testarmos a igualdade em cada um dos vetores da base $\{v_1, \ldots, v_m\} \subset F$. □

Corolário 4. *Se* $A\colon E \to F$ *é invertível então* $A^+ = A^{-1}$.

Evidente. □

Segue-se dos Corolários 2 e 3 que se A é injetiva então A^+ é uma inversa à esquerda de A, e se A é sobrejetiva então A^+ é uma inversa à direita de A.

Corolário 5. *Para toda transformação linear* $A\colon E \to F$*, tem-se* $(A^*)^+ = (A^+)^*\colon E \to F$.

Com efeito, as bases ortonormais fornecidas pelo Teorema 13.10 e usadas na definição de A' tanto servem para A como para A^*. Se as utilizarmos com A^* em vez de A, obteremos $(A^*)'u_i = (1/\sigma_i)v_i$ se $1 \leq i \leq r$ e $(A^*)'u_i = 0$ se $i \geq r+1$. Mas é claro que $(A')^*$ opera do mesmo modo sobre os vetores básicos u_i. Logo $(A')^* = (A^*)'$ e daí segue o corolário. □

Exemplo 16.1. Se $A\colon E \to F$ é ortogonal então é injetiva, logo $A^+ = (A^*A)^{-1}A^*$. Mas a ortogonalidade de A significa $A^*A = I_E$, logo $A^+ = A^*$.

Exemplo 16.2. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^3$ dada por $A(x,y) = (x,y,0)$. Como A é ortogonal, temos $A^+ = A^*$. A matriz de A tem colunas $(1,0,0)$ e $(0,1,0)$, logo estas são as linhas da matriz de A^*, portanto $A^*(x,y,z) = (x,y)$. Portanto a pseudo-inversa de A é $A^+\colon \mathbb{R}^3 \to \mathbb{R}^2$, dada por $A^+(x,y,z) = (x,y)$.

Exemplo 16.3. Definamos $A\colon \mathbb{R}^2 \to \mathbb{R}^3$ pondo $A(x,y) = (x,y,x+y)$. Como A é injetiva, sua pseudo-inversa é $A^+ = (A^*A)^{-1}A^*$. As colunas da matriz de A (linhas da matriz de A^*) são $(1,0,1)$ e $(0,1,1)$, logo

$$A^*(x,y,z) = (x+z, y+z)$$

e daí

$$A^*A(x,y) = (2x+y, x+2y).$$

Para determinar $(A^*A)^{-1}(x,y) = (s,t)$, resolvemos o sistema $(A^*A)(s,t) = (x,y)$, ou seja $2s+t = x$, $s+2t = y$, no qual as incógnitas

são s, t. Encontramos

$$s = \frac{(2x - y)}{3} \quad \text{e} \quad t = \frac{(2y - x)}{3}.$$

Portanto

$$(A^*A)^{-1}(x, y) = \frac{1}{3}(2x - y, 2y - x).$$

Assim, para qualquer $(x, y, z) \in \mathbb{R}^3$ temos

$$\begin{aligned} A^+(x, y, z) &= [(A^*A)^{-1} A^*](x, y, z) \\ &= (A^*A)^{-1}(x + z, y + z) \\ &= \frac{1}{3}(2x - y + z, 2y - x + z). \end{aligned}$$

Exemplo 16.4. Seja $B\colon \mathbb{R}^3 \to \mathbb{R}^2$ dada por

$$B(x, y, z) = \frac{1}{3}(2x - y + z, -x + 2y + z).$$

Temos as matrizes de posto 2:

$$\mathbf{b} = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{b}^T = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \\ 1 & 1 \end{bmatrix}.$$

Logo B é sobrejetiva e $B^+ = B^*(BB^*)^{-1}$. Como a matriz de B^* é $\mathbf{b}^T$, temos

$$B^*(x, y) = \frac{1}{3}(2x - y, -x + 2y, x + y)$$

para qualquer $(x, y) \in \mathbb{R}^2$. Segue-se que $BB^*\colon \mathbb{R}^2 \to \mathbb{R}^2$ é dado por

$$\begin{aligned} BB^*(x, y) &= \frac{1}{3}B(2x - y, -x + 2y, x + y) \\ &= \frac{1}{9}(6x - 3y, -3x + 6y) \\ &= \frac{1}{3}(2x - y, -x + 2y). \end{aligned}$$

Para determinar $(BB^*)^{-1}(x, y) = (s, t)$, resolvemos o sistema $BB^*(s, t) = (x, y)$, isto é, $2s - t = 3x$, $-s + 2t = 3y$, nas incógnitas s, t, e encontramos $s = 2x + y$, $t = x + 2y$, portanto

$$(BB^*)^{-1}(x, y) = (2x + y, x + 2y).$$

Isto nos dá, finalmente:

$$\begin{aligned} B^+(x,y) &= B^*(BB^*)^{-1}(x,y) \\ &= B^*(2x+y, x+2y) \\ &= \frac{1}{3}(3x, 3y, 3x+3y), \end{aligned}$$

ou seja, $B^+(x,y) = (x, y, x+y)$.
Retomando a transformação A do Exemplo 16.3, vemos que $B = A^+$ e constatamos que $(A^+)^+ = B^+ = A$.

A relação $A^{++} = A$, verificada no caso particular acima, é verdadeira em geral. Isto pode ser visto facilmente examinando a definição da transformação A' e notando que $(A')' = A$. Como $A' = A^+$, o resultado segue daí.

Exercícios

16.1. Determine a pseudo-inversa de cada uma das seguintes transformações lineares:

(a) A transformação nula $0\colon E \to F$;

(b) A projeção ortogonal $P\colon E \to E$ sobre o subespaço F;

(c) A mesma projeção acima, considerada como transformação linear de E sobre F;

(d) A projeção (não-ortogonal) $P\colon \mathbb{R}^2 \to \mathbb{R}^2$, sobre a reta F, paralelamente à reta G. (Descreva P^+ geometricamente.)

16.2. Para toda transformação linear $A\colon E \to F$ e todo $\alpha \neq 0$, prove que $(\alpha \cdot A)^+ = \frac{1}{\alpha} \cdot A^+$.

16.3. Identifique o núcleo e a imagem da pseudo-inversa de uma transformação linear $A\colon E \to F$.

16.4. Dada a transformação linear $A\colon E \to F$, prove que, para todo $w \in F$, tem-se $A^+AA^*w = A^*w$.

16.5. Sejam $A\colon E \to F$ e $B\colon F \to G$ transformações lineares. Se $\mathcal{I}m(A) = \mathcal{I}m(B^*)$, prove que $(BA)^+ = A^+B^+$.

16.6. Dado o operador $A\colon E \to E$, prove:

(a) Se A é auto-adjunto, A^+ também é.

(b) Se A é normal, A^+ também é.

(c) Se A é não-negativo, A^+ também é.

16.7. Dados os vetores linearmente independentes $v_1, \ldots, v_r \in \mathbb{R}^n$, seja $\mathbf{a} \in M(n \times r)$ a matriz que os tem como colunas. Prove que a projeção ortogonal de um vetor qualquer $x \in \mathbb{R}^n$ sobre o subespaço gerado por $v_1, \ldots, v_r$ é $Px = \mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}\,\mathbf{a}^T\mathbf{x}$, onde identificamos o vetor $x \in \mathbb{R}^n$ com a matriz $\mathbf{x} \in M(n \times 1)$ cuja única coluna é x.

16.8. Use a fórmula acima para determinar a projeção ortogonal do vetor $(1, 2, 3)$ sobre o plano gerado pelos vetores $(1, 1, 1)$ e $(1, -1, 1)$.

16.9. Seja $A\colon \mathbb{R}^n \to \mathbb{R}^m$ uma transformação linear. Prove que, dado qualquer $b \in \mathbb{R}^m$, a equação $A^*Ax = A^*b$ sempre possui solução. (Uma infinidade delas se A^*A não é invertível.)

16.10. Prove as seguintes propriedades da pseudo-inversa de uma transformação linear $A\colon E \to F$:

(a) $AA^+A = A$

(b) $A^+AA^+ = A^+$

(c) $(AA^+)^* = AA^+$

(d) $(A^+A)^* = A^+A$.

16.11. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^3$ dada por $A(x, y) = (x, y, 2x + 3y)$. Determine a pseudo-inversa $A^+\colon \mathbb{R}^3 \to \mathbb{R}^2$.

16.12. Ache a pseudo-inversa da transformação linear $A\colon \mathbb{R}^3 \to \mathbb{R}^2$, sabendo que $A(x, y, z) = (x + y, y + z)$.

16.13. Determine a pseudo-inversa de uma matriz diagonal $\mathbf{d} = [d_{ij}] \in M(n \times n)$. (Isto significa, naturalmente, a matriz de D^+, onde $D\colon \mathbb{R}^n \to \mathbb{R}^n$ é o operador linear cuja matriz na base canônica é $\mathbf{d}$.) Considere explicitamente o caso em que a diagonal de $\mathbf{d}$ é $(1, 0, -2, 0, 1/3)$.

16.14. Dada a transformação linear (não necessariamente injetiva) $A\colon E \to F$, sejam $P\colon E \to E$ e $Q\colon F \to F$ as projeções ortogonais sobre $\mathrm{Im}(A^*)$ e $\mathrm{Im}(A)$ respectivamente. Interprete e demonstre a igualdade $A^+ = PA^{-1}Q$.

16.15. Defina o operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ pondo $A(x, y) = (x - y, x - y)$. Determine a matriz de $A^+\colon \mathbb{R}^2 \to \mathbb{R}^2$ na base canônica. Ache o vetor $v \in \mathbb{R}^2$ de menor norma tal que Av está o mais próximo possível de $w = (3, 5)$.

16.16. Dado o vetor não-nulo $a \in E$, defina a transformação linear $A\colon \mathbb{R} \to E$ pondo $A \cdot 1 = a$. Mostre que $A^*\colon E \to \mathbb{R}$ é definida por $A^* \cdot w = \langle a, w \rangle$ e, usando a expressão $A^+ = (A^*A)^{-1} A^*$, conclua que $A^+\colon E \to \mathbb{R}$ é dada por $A^+ \cdot w = \langle a, w \rangle / |a|^2$.

16.17. Fixado o vetor não-nulo $b \in E$, defina $B\colon E \to \mathbb{R}$ pondo $B \cdot w = \langle b, w \rangle$. Mostre que $B^+ = B^*(BB^*)^{-1}$ para concluir que $B^+\colon \mathbb{R} \to E$ cumpre $B^+ \cdot 1 = b/|b|^2$.

16.18. Sejam $A\colon \mathbb{R} \to E$ e $B\colon E \to \mathbb{R}$ definidas por $A \cdot 1 = a$, $B \cdot w = \langle b, w \rangle$, onde $a, b \in E$ são vetores fixados de tal modo que $\langle a, b \rangle \neq 0$. Prove que as transformações lineares $(BA)^+\colon \mathbb{R} \to \mathbb{R}$ e $A^+B^+\colon \mathbb{R} \to \mathbb{R}$ são dadas por

$$(BA)^+ \cdot 1 = \frac{1}{\langle a, b \rangle} \quad \text{e} \quad (A^+B^+) \cdot 1 = \frac{\langle a, b \rangle}{|a|^2 \cdot |b|^2}.$$

Conclua que, em geral, se tem $(BA)^+ \neq A^+B^+$. Dê um exemplo concreto desta desigualdade, com $A\colon \mathbb{R} \to \mathbb{R}^2$ e $B\colon \mathbb{R}^2 \to \mathbb{R}$.

16.19. Com a notação dos dois exercícios anteriores, prove que $(AB)^+ = B^+A^+$.

16.20. Seja $A\colon E \to F$ a transformação linear de posto 1 dada por $Av = \langle v, a\rangle\, b$, onde $a \in E$ e $b \in F$ são vetores $\neq 0$. Sabendo que $A^*w = \langle w, b\rangle\, a$ para todo $w \in F$ (cfr. Exercício 11.17), mostre que

$$A^+w = \frac{\langle w, b\rangle}{|a|^2|b|^2} \cdot a$$

para todo $w \in F$.

16.21. Sejam $A\colon \mathbb{R}^m \to \mathbb{R}^n$ sobrejetiva e $B\colon \mathbb{R}^n \to \mathbb{R}^m$ injetiva. Prove que $(BA)^+ = A^+B^+$. (Generalização do Exercício 16.19.)

16.22. Prove que a projeção $P\colon E \to E$ é ortogonal se, e somente se, $P^+ = P$.

16.23. Use o Exercício 16.20 para calcular a pseudo-inversa da transformação linear $A\colon \mathbb{R}^2 \to \mathbb{R}^3$ que tem a matriz

$$\mathbf{a} = \begin{bmatrix} 1 & 3 \\ 4 & 12 \\ 1 & 3 \end{bmatrix}.$$

16.24. Sejam $v_1, v_2, v_3, v_4, v_5 \in \mathbb{R}^4$ as colunas de uma matriz $[a_{ij}] \in M(4 \times 5)$. Suponha que v_1 e v_2 sejam L.I. e que $v_3 = b_{13}v_1 + b_{23}v_2$, $v_4 = b_{14}v_1 + b_{24}v_2$, $v_5 = b_{15}v_1 + b_{25}v_2$. Prove que

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \begin{bmatrix} 1 & 0 & b_{13} & b_{14} & b_{15} \\ 0 & 1 & b_{23} & b_{24} & b_{25} \end{bmatrix}.$$

Mostre que este é um método geral para exprimir toda matriz $m \times n$ de posto r como produto de uma matriz $m \times r$ por uma matriz $r \times n$, ambas de posto (máximo) igual a r. Compare com o Exercício 6.29, do qual esta é uma versão matricial. Mostre que as matrizes 4×2 e 2×5 acima foram obtidas a partir da solução natural daquele exercício. Deduza daí (com auxílio do Exercício 16.21) um processo para calcular a pseudo-inversa de qualquer matriz.

17

Tópicos Matriciais

Salvo o item C e a observação final do item G, os assuntos tratados nesta seção não serão necessários para entender as seguintes. Alguns deles são traduções, para a linguagem das matrizes, de teoremas e métodos apresentados nas seções precedentes, outros são tópicos matriciais interessantes em si mesmos ou temas clássicos do cálculo matricial que se têm revelado úteis, especialmente sob o ponto de vista computacional.

17.A Matrizes de Gram

Seja E um espaço vetorial de dimensão finita, munido de produto interno. A *matriz de Gram* dos vetores $v_1, \ldots, v_k \in E$ é a matriz $\mathbf{g} = [g_{ij}] \in M(k \times k)$, onde $g_{ij} = \langle v_i, v_j \rangle$. Quando precisarmos ser mais explícitos, escreveremos $\mathbf{g} = \mathbf{g}(v_1, \ldots, v_k)$.

Dada uma base $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$, seja $\mathbf{a} = [a_{ij}] \in M(n \times k)$ a matriz das coordenadas dos vetores v_j em relação à base $\mathcal{U}$, isto é:

$$v_j = a_{1j}u_1 + \cdots + a_{nj}u_n \quad \text{para} \quad j = 1, \ldots, k.$$

Seja ainda $\mathbf{h} = [h_{ij}] \in M(n \times n)$ a matriz de Gram da base $\mathcal{U}$, isto é, $h_{ij} = \langle u_i, u_j \rangle$. Então, para $i, j = 1, \ldots, k$, temos (escrevendo $\mathbf{m}_{ij}$ para

indicar o ij-ésimo elemento de uma matriz $\mathbf{m}$):

$$
\begin{aligned}
g_{ij} = \langle v_i, v_j \rangle &= \left\langle \sum_{r=1}^{n} a_{ri} u_r, \sum_{s=1}^{n} a_{sj} u_s \right\rangle = \sum_{r,s=1}^{n} a_{ri} a_{sj} h_{rs} \\
&= \sum_{r=1}^{n} a_{ri} \left(\sum_{s=1}^{n} h_{rs} a_{sj} \right) = \sum_{r=1}^{n} (\mathbf{a}^T)_{ir} (\mathbf{ha})_{rj} = (\mathbf{a}^T \mathbf{ha})_{ij}
\end{aligned}
$$

portanto $\mathbf{g} = \mathbf{a}^T \mathbf{ha}$.

Em particular, se tomarmos uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$, teremos $\mathbf{h} = \mathbf{I}_n$, portanto a matriz de Gram $\mathbf{g}$ se escreve como

$$\mathbf{g} = \mathbf{g}(v_1, \ldots, v_k) = \mathbf{a}^T \cdot \mathbf{a},$$

onde $\mathbf{a}$ é a matriz das coordenadas dos vetores v_j em relação a uma base ortonormal de E. Daí resultam:

1) *Toda matriz de Gram é não-negativa*;

2) *A matriz de Gram $\mathbf{g} = \mathbf{g}(v_1, \ldots, v_k)$ é positiva (isto é, invertível) se, e somente se, os vetores $v_1, \ldots, v_k$ são* L.I..

Reciprocamente, se uma matriz $\mathbf{g} = [g_{ij}] \in M(k \times k)$ admite uma decomposição do tipo $\mathbf{g} = \mathbf{a}^T \cdot \mathbf{a}$, onde $\mathbf{a} = [a_{ij}] \in M(n \times k)$ então, tomando uma base ortonormal $\{u_1, \ldots, u_n\} \subset E$ e escrevendo

$$v_j = \sum_{i=1}^{n} a_{ij} u_i \quad (j = 1, \ldots, k),$$

obtemos vetores $v_1, \ldots, v_k \in E$ tais que

$$\langle v_i, v_j \rangle = \sum_{k=1}^{n} a_{ki} a_{kj} = (\mathbf{a}^T \cdot \mathbf{a})_{ij} = g_{ij}\,,$$

logo $\mathbf{g} = \mathbf{g}(v_1, \ldots, v_k)$ é a matriz de Gram dos vetores $v_1, \ldots, v_k$. Isto leva a mais uma propriedade das matrizes de Gram:

3) *Toda matriz não-negativa $\mathbf{g} = [g_{ij}] \in M(k \times k)$ é a matriz de Gram de uma lista de vetores $v_1, \ldots, v_k \in E$.*

Com efeito, existe $\mathbf{a} \in M(k \times k)$ simétrica (e não-negativa) tal que $\mathbf{g} = \mathbf{a}^2 = \mathbf{a}^T \cdot \mathbf{a}$ (Teorema 13.8).

17.B Matrizes Triangulares

Uma matriz $\mathbf{t} = [t_{ij}] \in M(n \times n)$ diz-se *triangular superior* quando $t_{ij} = 0$ para $i > j$ e *triangular inferior* quando $t_{ij} = 0$ para $i < j$.

Trataremos primordialmente de matrizes triangulares superiores. As propriedades das triangulares inferiores são análogas e se provam analogamente.

Uma matriz triangular superior é a matriz de um operador linear $T\colon \mathbb{R}^n \to \mathbb{R}^n$ tal que $\langle e_i, Te_j \rangle = 0$ para $i > j$. Se chamarmos de $F_i \subset \mathbb{R}^n$ o subespaço vetorial formado pelos vetores $(x_1, \ldots, x_i, \ldots, 0)$ cujas últimas $n - i$ coordenadas são nulas, a matriz do operador $T\colon \mathbb{R}^n \to \mathbb{R}^n$ (na base canônica) é triangular superior se, e somente se, todos os subespaços $F_o \subset \cdots \subset F_n$ são invariantes por T.

Com efeito, a condição $\langle e_i, Te_j \rangle = 0$ para $i > j$ significa que, para cada j, as últimas $n - j$ coordenadas do vetor Te_j são nulas, ou seja, que $Te_j \in F_j$ para $j = 1, \ldots, n$. Isto é o mesmo que dizer que $T(F_j) \subset F_j$ para todo j.

Seguem-se algumas propriedades das matrizes triangulares superiores:

1) *O produto de duas matrizes triangulares superiores é ainda uma matriz triangular superior.*

Com efeito, se o subespaço $F_i \subset \mathbb{R}^n$ é invariante por cada um dos operadores $S, T\colon \mathbb{R}^n \to \mathbb{R}^n$ então F_i é invariante pelo produto ST.

2) *Uma matriz triangular superior $\mathbf{t} = [t_{ij}]$ é invertível se, e somente se, os elementos t_{ii} da sua diagonal são todos diferentes de zero. No caso afirmativo, a inversa $\mathbf{t}^{-1}$ é triangular superior e os elementos de sua diagonal são t_{ii}^{-1}.*

Com efeito, se todos os t_{ii} são $\neq 0$, dado um vetor não-nulo $v \in \mathbb{R}^n$ provemos que $Tv \neq 0$ (onde T é o operador de $\mathbb{R}^n$ cuja matriz na base canônica é $\mathbf{t}$). Seja $v = x_1 e_1 + \cdots + x_r e_r$, com $x_r \neq 0$. Como $\langle e_r, Te_i \rangle = 0$ para $i < r$, temos

$$\langle e_r, Tv \rangle = \sum_{i \leq r} \langle e_r, x_i Te_i \rangle = x_r \langle e_r, Te_r \rangle = x_r t_{rr},$$

portanto $\langle e_r, Tv \rangle \neq 0$, e daí $Tv \neq 0$. Assim $\mathbf{t}$ é invertível.

Reciprocamente, se $\mathbf{t}$ (ou seja, T) é invertível então, para cada $i = 1, \ldots, n$, a restrição $T\colon F_i \to F_i$, de T ao subespaço F_i, é também

invertível, logo sobrejetiva. Se fosse $t_{ii} = \langle e_i, Te_i \rangle = 0$, teríamos, como acabamos de ver, para todo $v = x_1 e_1 + \cdots + x_i e_i \in F_i$, $\langle e_i, Tv \rangle = x_i t_{ii} = 0$, logo $Tv \in F_{i-1}$. Isto significa que $T(F_i) \subset F_{i-1}$, contradizendo a sobrejetividade de $T \colon F_i \to F_i$. Portanto todos os t_{ii} são diferentes de zero.

Se a matriz triangular superior $\mathbf{t}$ é invertível, o operador linear $T \colon \mathbb{R}^n \to \mathbb{R}^n$ também é. Cada um dos subespaços $F_i = S(e_1, \ldots, e_i) \subset \mathbb{R}^n$ sendo invariante por T é também invariante por T^{-1}. (Com efeito, $A(F) \subset F \Rightarrow A(F) = F \Rightarrow A^{-1}(F) = F$.) Logo a matriz $\mathbf{t}^{-1}$, do operador T^{-1}, é também triangular superior. Escrevendo $\mathbf{t}^{-1} = [s_{ij}]$, a igualdade $\mathbf{t}^{-1}t = \mathbf{I}_n$ nos dá, para cada $i = 1, \ldots, n$:

$$1 = (\mathbf{t}^{-1}\mathbf{t})_{ii} = \sum_{k=1}^{n} s_{ik} t_{ki} = s_{ii} t_{ii},$$

pois $s_{ik} = 0$ se $i > k$ e $t_{ki} = 0$ se $k > i$. Logo $s_{ii} = 1/t_{ii}$.

3) *Os autovalores de uma matriz triangular superior* $\mathbf{t} = [t_{ij}] \in M(n \times n)$ *são os elementos* t_{ii} *da sua diagonal.*

Por definição, os autovalores de $\mathbf{t}$ são aqueles do operador $T \colon \mathbb{R}^n \to \mathbb{R}^n$ cuja matriz na base canônica é $\mathbf{t}$.

Em primeiro lugar, se λ é um autovalor de $\mathbf{t}$, isto é, se existe $n \neq 0$ em $\mathbb{R}^n$ tal que $Tv = \lambda \cdot v$, seja $v = x_1 e_1 + \cdots + x_r e_r$, com $x_r \neq 0$. Então, como $\langle e_r, Te_i \rangle = 0$ se $i < r$, temos

$$\begin{aligned} \lambda x_r = \langle e_r, \lambda v \rangle &= \langle e_r, Tv \rangle \\ &= \langle e_r, x_1 Te_1 + \cdots + x_r Te_r \rangle \\ &= \langle e_r, x_r Te_r \rangle = t_{rr} x_r. \end{aligned}$$

Segue-se que $\lambda = t_{rr}$. Assim, somente os números t_{ii} podem ser autovalores de T.

Em segundo lugar, todos os t_{ii} são, de fato, autovalores de $\mathbf{t}$. Com efeito, a matriz $\mathbf{t} - t_{ii}\mathbf{I}_n$ é triangular superior e o i-ésimo elemento de sua diagonal é zero, logo não é invertível. Conseqüentemente, t_{ii} é um autovalor de $\mathbf{t}$.

4) *Seja* $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ *uma base ortonormal, obtida pelo processo de Gram-Schmidt a partir da base* $\mathcal{V} = \{v_1, \ldots, v_n\} \subset E$. *A matriz de passagem de* $\mathcal{V}$ *para* $\mathcal{U}$ *é triangular superior e seus autovalores são todos positivos.*

Com efeito, escrevendo

$$u_j = \sum_{i=1}^{n} p_{ij} v_i,$$

sabemos que cada u_j pertence ao subespaço vetorial gerado por $v_1, \ldots, v_j$, logo $u_j = p_{1j}v_1 + \cdots + p_{jj}v_j$. Isto mostra que a matriz de passagem $\mathbf{p} = [p_{ij}] \in M(n \times n)$ é triangular superior. Além disso, como $u_j = |w_j|^{-1}\, w_j$, onde o vetor w_j tem a forma

$$w_j = v_j - \sum_{i<j} \alpha_{ij} v_i,$$

vemos que $p_{jj} = |w_j|^{-1}$ é positivo, para cada $j = 1, \ldots, n$. Como vimos acima, esses números p_{jj} são os autovalores da matriz $\mathbf{p}$.

17.C Decomposição de Cholesky

Mostraremos que toda matriz positiva $\mathbf{a} = [a_{ij}] \in M(n \times n)$ pode ser expressa como o produto $\mathbf{a} = \mathbf{t}^T \cdot \mathbf{t}$, onde $\mathbf{t} \in M(n \times n)$ é uma matriz triangular superior cujos elementos da diagonal são todos positivos. A expressão $\mathbf{a} = \mathbf{t}^T \cdot \mathbf{t}$ chama-se a *decomposição de Cholesky* da matriz $\mathbf{a}$.

Mais adiante (17.G) daremos outra prova da existência da decomposição de Cholesky, por um método que permite obter a matriz $\mathbf{t}$ por escalonamento a partir de $\mathbf{a}$.

Agora, demonstraremos a possibilidade da decomposição $\mathbf{a} = \mathbf{t}^T \cdot \mathbf{t}$ usando Gram-Schmidt e a existência da raiz quadrada de $\mathbf{a}$ quando $\mathbf{a} > 0$ (isto é, quando $\mathbf{a}$ é positiva).

Como vimos acima (17.A) existem vetores $v_1, \ldots, v_n \in \mathbb{R}^n$ tais que $a_{ij} = \langle v_i, v_j \rangle$, ou seja, $\mathbf{a} = \mathbf{g}(v_1, \ldots, v_n)$ é a matriz de Gram dos vetores $v_1, \ldots, v_n$, os quais formam uma base de $\mathbb{R}^n$, pois $\mathbf{a} > 0$.

Pelo processo de Gram-Schmidt, obtemos uma base ortonormal $\{u_1, \ldots, u_n\} \subset \mathbb{R}^n$ a partir de $v_1, \ldots, v_n$. Para $i, j = 1, \ldots, n$, temos

$$u_i = \sum_r p_{ri} v_r, \quad u_j = \sum_s p_{sj} v_s,$$

onde a matriz de passagem $\mathbf{p} = [p_{ij}]$ é triangular superior, com $p_{ii} > 0$

para todo i. (17.B.) Usando o símbolo de Kronecker δ_{ij}, temos

$$\delta_{ij} = \langle u_i, u_j \rangle = \sum_{r,s=1}^{n} p_{ri} p_{sj} \langle v_r, v_s \rangle = \sum_{r,s=1}^{n} p_{ri} a_{rs} p_{sj} \,,$$

logo $\mathbf{p}^T \mathbf{a} \mathbf{p} = \mathbf{I}_n$. Pondo $\mathbf{t} = \mathbf{p}^{-1}$, obtemos $\mathbf{a} = \mathbf{t}^T \cdot \mathbf{t}$.

A decomposição de Cholesky é única. Noutras palavras, se $\mathbf{s}$ e $\mathbf{t}$ são matrizes triangulares superiores $n \times n$ com diagonais positivas e $\mathbf{s}^T \cdot \mathbf{s} = \mathbf{t}^T \cdot \mathbf{t}$ então $\mathbf{s} = \mathbf{t}$.

Com efeito, de $\mathbf{s}^T.\mathbf{s} = \mathbf{t}^T.\mathbf{t}$ resulta $\mathbf{s}\mathbf{t}^{-1} = (\mathbf{s}^T)^{-1}.\mathbf{t}^T$. Como o primeiro membro desta última igualdade é uma matriz triangular superior e o segundo é triangular inferior, concluímos que $\mathbf{d} = \mathbf{s}\mathbf{t}^{-1} = (\mathbf{s}^T)^{-1}.\mathbf{t}^T$ é uma matriz diagonal, com $d_{ii} > 0$ (e $d_{ij} = 0$ se $i \neq j$). Segue-se imediatamente das igualdades acima que $\mathbf{s} = \mathbf{d}\mathbf{t}$ e $\mathbf{t} = \mathbf{d}\mathbf{s}$. Olhando para os elementos da diagonal, temos $s_{ii} = d_{ii} t_{ii}$ e $t_{ii} = d_{ii} s_{ii}$. Como $s_{ii} > 0$ e $t_{ii} > 0$, isto implica $d_{ii} = 1$, logo $\mathbf{d} = \mathbf{I}_n$ e $\mathbf{s} = \mathbf{t}$.

17.D A Decomposição qr

Esta é uma interpretação matricial do processo de Gram-Schmidt. Segundo ela, *toda matriz invertível* $\mathbf{a} = [a_{ij}] \in M(n \times n)$ *admite uma decomposição do tipo* $\mathbf{a} = \mathbf{q}\mathbf{r}$, *onde* $\mathbf{q}$ *é ortogonal e* $\mathbf{r}$ *é triangular superior, com elementos positivos na diagonal*.

Para chegar a este resultado, chamemos de $v_1, \ldots, v_n$ as colunas da matriz $\mathbf{a}$ e de $\mathcal{U} = \{u_1, \ldots, u_n\} \subset \mathbb{R}^n$ a base ortonormal obtida dos v_i pelo processo de Gram-Schmidt. Como sabemos, a matriz $\mathbf{p} = [p_{ij}]$ de passagem da base $\mathcal{V} = \{v_1, \ldots, v_n\}$ para a base $\mathcal{U}$ é triangular superior, com elementos positivos na diagonal. Além disso, a matriz $\mathbf{q} = [q_{ij}]$, cujas colunas são os vetores $u_j = (q_{1j}, q_{2j}, \ldots, q_{nj})$, é ortogonal.

Tomando a i-ésima coordenada de ambos os membros da igualdade vetorial $u_j = \sum\limits_k p_{kj} v_k$, obtemos

$$q_{ij} = \sum_{k=1}^{n} p_{kj} a_{ik} = \sum_{k=1}^{n} a_{ik} p_{kj} = (\mathbf{a}\mathbf{p})_{ij} \,,$$

para quaisquer $i, j = 1, \ldots, n$. Logo $\mathbf{q} = \mathbf{ap}$. A matriz $\mathbf{r} = \mathbf{p}^{-1}$ é, como vimos acima, (17.B), triangular superior com elementos positivos na diagonal. De $\mathbf{ap} = \mathbf{q}$ resulta imediatamente $\mathbf{a} = \mathbf{qr}$.

Observação Dada a matriz invertível $\mathbf{a} \in M(n \times n)$, são únicas as matrizes $\mathbf{q}$, $\mathbf{r}$ tais que $\mathbf{a} = \mathbf{qr}$, $\mathbf{q}$ é ortogonal, $\mathbf{r}$ é triangular superior e os elementos de sua diagonal são positivos. Com efeito, $\mathbf{a} = \mathbf{qr}$ escreve-se também como $\mathbf{ap} = \mathbf{q}$, onde $\mathbf{p} = \mathbf{r}^{-1}$. Isto quer dizer que $\mathbf{p}$ é a matriz de passagem da base $\mathcal{V} = \{v_1, \ldots, v_n\} \subset \mathbb{R}^n$, formada pelas colunas de $\mathbf{a}$, para a base ortonormal $\mathcal{U} = \{u_1, \ldots, u_n\} \subset \mathbb{R}^n$, dada pelas colunas de $\mathbf{q}$. Ora, as quatro condições seguintes implicam que cada u_j é determinado univocamente a partir de $v_1, \ldots, v_j$:

1) $u_j = p_{1j}v_1 + p_{2j}v_2 + \cdots + p_{jj}v_j$;

2) u_j é ortogonal a $v_1, \ldots, v_{j-1}$;

3) $|u_j| = 1$;

4) $p_{jj} > 0$.

Com efeito, 1) diz que u_j pertence ao subespaço $F \subset \mathbb{R}^n$ gerado por $v_1, \ldots, v_j$. 2) diz que u_j pertence a reta R, complemento ortogonal de $\{v_1, \ldots, v_{j-1}\}$ no subespaço F. Já 3) restringe u_j a ser um dos 2 vetores unitários da reta R. E, finalmente, 4) diz que u_j é o vetor unitário de R tal que $\langle u_j, v_j \rangle$ é positivo.

A condição 1) diz que a matriz $\mathbf{p}$ é triangular superior. 2) e 3) dizem que a matriz $\mathbf{q}$ é ortogonal, enquanto 4) afirma que os elementos da diagonal de $\mathbf{p}$ são positivos. Juntas, elas garantem a unicidade de $\mathbf{q}$ e portanto a unicidade de $\mathbf{p} = \mathbf{a}^{-1}\mathbf{q}$.

A observação acima estabelece também a unicidade do processo de Gram-Schmidt sob a condição de que cada u_j pertença ao subespaço gerado por $v_1, \ldots, v_j$ e cumpra $\langle u_j, v_j \rangle > 0$ (além, naturalmente, de serem $u_1, \ldots, u_j$ ortonormais).

Em resumo: a igualdade $\mathbf{a} = \mathbf{qr}$ significa que as colunas de $\mathbf{q}$ formam a base ortonormal de $\mathbb{R}^n$ obtida das colunas de $\mathbf{a}$ por Gram-Schmidt e $\mathbf{r}$ é a matriz de passagem das colunas de $\mathbf{q}$ para as colunas de $\mathbf{a}$.

17.E Diagonalização, Decomposição Polar e Valores Singulares

O Corolário do Teorema 12.2 e o Teorema Espectral 13.6, quando formulados em termos de matrizes, apresentam as seguintes versões:

1) *Se a matriz quadrada* $\mathbf{a} \in M(n \times n)$ *possui* n *autovalores diferentes então existe uma matriz invertível* $\mathbf{p} \in M(n \times n)$ *tal que* $\mathbf{p}^{-1}\mathbf{a}\mathbf{p} = \mathbf{d}$ *é uma matriz diagonal. Os elementos da diagonal de* $\mathbf{d}$ *são os autovalores de* $\mathbf{a}$.

2) *Se a matriz* $\mathbf{a} \in M(n \times n)$ *é simétrica então existe uma matriz ortogonal* $\mathbf{q} \in M(n \times n)$ *tal que* $\mathbf{q}^T\mathbf{a}\mathbf{q} = \mathbf{d}$ *é uma matriz diagonal. Os elementos da diagonal de* $\mathbf{d}$ *são os autovalores de* $\mathbf{a}$.

Já o Teorema 14.4 assume a seguinte forma:

3) *Toda matriz quadrada* $\mathbf{a} \in M(n \times n)$ *se exprime como produto* $\mathbf{a} = \mathbf{p}\mathbf{u}$, *onde* $\mathbf{p} \in M(n \times n)$ *é uma matriz não-negativa e* $\mathbf{u}$ *é ortogonal*.

Finalmente, o Teorema 13.10 tem a versão matricial abaixo:

4) *Para toda matriz* $\mathbf{a} \in M(m \times n)$ *existem matrizes ortogonais* $\mathbf{p} \in M(m \times m)$, $\mathbf{q} \in M(n \times n)$, *tais que* $\mathbf{p}\mathbf{a}\mathbf{q} = \mathbf{d}$, *onde* $\mathbf{d} \in M(m \times n)$ *é uma matriz diagonal,* $\mathbf{d} = [d_{ij}]$, *isto é,* $d_{ij} = 0$ *se* $i \neq j$. *Para* $i = 1, \ldots, r =$ *posto de* $\mathbf{a}$, *tem-se* $d_{ii} = \sigma_i > 0$, σ_i^2 *é um autovalor de* $\mathbf{a}^T\mathbf{a}$ *(e de* $\mathbf{a}\mathbf{a}^T$*) e, para* $i > r$, $d_{ii} = 0$.

Equivalentemente, podemos escrever $\mathbf{a} = \mathbf{p}^T\mathbf{d}\mathbf{q}^T$ ou, mudando de notação, $\mathbf{a} = \mathbf{p}\mathbf{d}\mathbf{q}$. Esta se chama a *decomposição de* $\mathbf{a}$ *a valores singulares* (singular value decomposition), pois os elementos não-nulos σ_i da diagonal de $\mathbf{d}$ são os valores singulares de $\mathbf{a}$, ou seja, $\sigma_i > 0$ para todo $i = 1, \ldots, n$ e, além disso,

$$\sigma_1^2, \ldots, \sigma_r^2$$

são autovalores de $\mathbf{a}^T\mathbf{a}$.

Se $\mathbf{a} \in M(4 \times 5)$ e tem posto 3 então a matriz $\mathbf{d}$ tem a seguinte forma, com $\lambda_1 > 0$, $\lambda_2 > 0$ e $\lambda_3 > 0$:

$$\mathbf{d} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

17.F A Decomposição lu

Cada uma das três operações elementares sobre as linhas de uma matriz, introduzidas na Seção 9, pode ser interpretada como a multiplicação à esquerda por uma matriz invertível de tipo especial, chamada uma matriz elementar.

Mais precisamente, uma *matriz elementar* $m \times m$ é uma matriz que resulta da aplicação de uma operação elementar à matriz identidade $\mathbf{I}_m$. Há portanto 3 tipos de matrizes elementares. Vejamos alguns exemplos no caso 4×4:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \alpha & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

As matrizes acima se obtêm a partir de $\mathbf{I}_4$ mediante as operações $L_1 \leftrightarrow L_3$, $L_4 + \alpha L_1$ e αL_2 respectivamente.

Afirmamos que aplicar uma operação elementar a uma matriz com m linhas é o mesmo que multiplicá-la à esquerda pela matriz que resulta de aplicar a mesma operação às linhas de $\mathbf{I}_m$.

Isto decorre da seguinte observação (cuja verificação deixamos a cargo do leitor): se $\mathbf{a} \mapsto \mathbf{a}'$ simboliza uma determinada operação elementar sobre matrizes com m linhas então, para toda $\mathbf{b} \in M(m \times m)$, tem-se $(\mathbf{ba})' = \mathbf{b}' \cdot \mathbf{a}$. Admitido este fato, tomamos $\mathbf{m} = \mathbf{I}_m'$ e vemos que, para toda matriz $\mathbf{a}$ com m linhas, vale

$$\mathbf{a}' = (\mathbf{I}_m \cdot \mathbf{a})' = \mathbf{I}_m' \cdot \mathbf{a} = \mathbf{m} \cdot \mathbf{a}.$$

Portanto o método de eliminação de Gauss para reduzir uma matriz de m linhas à forma escalonada consiste em multiplicá-la sucessivamente à esquerda por matrizes elementares do tipo 1 (transposição de duas linhas) ou do tipo 2 (subtrair de uma linha um múltiplo de outra linha). Uma matriz elementar do tipo 2, que corresponda à

operação $L_i - \alpha L_j$, com $j < i$, tem a forma abaixo:

$$\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \vdots & \ddots & & & \\ & & -\alpha & \dots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

Os elementos não indicados fora da diagonal são nulos. O número $-\alpha$ está na linha i e na coluna j. A fim de tornar igual a zero o elemento a_{ij} da matriz $\mathbf{a} = [a_{ij}]$ mediante essa pré-multiplicação, deve-se tomar $\alpha = a_{ij}/a_{jj}$. O elemento a_{jj}, que se supõe $\neq 0$, chama-se o *pivô*. Para eliminar (tornar iguais a zero) todos os elementos abaixo da diagonal na coluna j da matriz $\mathbf{a}$, deve-se pré-multiplicá-la pelo produto das $m-j$ matrizes da forma acima que se obtêm fazendo sucessivamente $i = j+1, \ldots, m$. Isto equivale a pré-multiplicar pela matriz

$$\mathbf{m}_j = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & -\alpha_{j+1,j} & & \\ & & \vdots & \ddots & \\ & & -\alpha_{mj} & \dots & 1 \end{bmatrix} \qquad (*)$$

(com os α's na j-ésima coluna), onde $\alpha_{rj} = a_{rj}/a_{jj}$. Por exemplo, se $m = 4$ e $j = 1$, vemos facilmente que

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\alpha_2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\alpha_3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\alpha_4 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\alpha_2 & 1 & 0 & 0 \\ -\alpha_3 & 0 & 1 & 0 \\ -\alpha_4 & 0 & 0 & 1 \end{bmatrix}.$$

Suponhamos que, durante o processo de escalonamento de uma dada matriz, nunca haja necessidade de se efetuar transposição de linhas. Então podemos assegurar que existem matrizes $\mathbf{m}_1, \ldots, \mathbf{m}_m$ do tipo (*) acima, tais que

$$\mathbf{m}_m \ldots \mathbf{m}_2 \mathbf{m}_1 \mathbf{a} = \mathbf{u},$$

onde $\mathbf{u}$ é uma matriz escalonada. Se $\mathbf{a}$ for uma matriz quadrada, $\mathbf{u}$ é triangular superior. (Daí a notação: $\mathbf{u}$ provém de upper triangular, em inglês.) Se $\mathbf{u} = [u_{ij}] \in M(m \times n)$ (além de não requerer transposições de linha em seu escalonamento) tem posto máximo (m ou n) então o primeiro elemento não-nulo de sua i-ésima linha é u_{ii}. Portanto, neste caso, os elementos u_{ii} da diagonal de $\mathbf{u}$ são os pivôs, logo são diferentes de zero.

Evidentemente, toda matriz elementar é invertível, logo toda matriz $\mathbf{m}_j$ do tipo (*) acima possui uma inversa e é claro que

$$\mathbf{m}_j^{-1} = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & \alpha_{j+1,j} & & \\ & & \vdots & \ddots & \\ & & \alpha_{mj} & & 1 \end{bmatrix}$$

Segue-se então que toda matriz $\mathbf{a}$ cujo escalonamento não requer transposições de linhas se escreve como

$$\mathbf{a} = \mathbf{m}_1^{-1} \mathbf{m}_2^{-1} \ldots \mathbf{m}_m^{-1}\, \mathbf{u} = \mathbf{l}\mathbf{u}$$

onde as $\mathbf{m}_j$ têm a forma (*) e $\mathbf{u}$ é escalonada.

Agora ocorre um fato notável. Embora o produto $\mathbf{m}_m \ldots \mathbf{m}_2 \mathbf{m}_1$ não tenha nenhuma expressão especial, vale a igualdade

$$\mathbf{l} = \mathbf{m}_1^{-1} \mathbf{m}_2^{-1} \ldots \mathbf{m}_m^{-1} = \begin{bmatrix} 1 & & & & \\ \alpha_{21} & 1 & & & \\ \alpha_{31} & \alpha_{32} & \ddots & & \\ \vdots & \vdots & & 1 & \\ \alpha_{m1} & \alpha_{m2} & & \alpha_{m,m-1} & 1 \end{bmatrix},$$

onde $\alpha_{ij} = a_{ij}^{(j)}/a_{jj}^{(j)}$. Nesta notação, indicamos com $\mathbf{a}^{(j)} = \mathbf{m}_j \ldots \mathbf{m}_1$ a matriz que se obtém depois da eliminação dos elementos abaixo da diagonal na j-ésima coluna. (Portanto $j = 0, 1, \ldots, m-1$, e $\mathbf{a}^{(0)} = \mathbf{a}$.) Estamos admitindo que não houve necessidade de efetuar transposição de linhas, ou seja, que em todas as etapas se tem o pivô $(\mathbf{a}^{(j)})_{jj} \neq 0$.

Concluímos então que, mediante essa hipótese, a matriz **a** se escreve como um produto

$$\mathbf{a} = \mathbf{lu},$$

onde $\mathbf{l} \in M(m \times m)$ é uma matriz triangular inferior (lower triangular) com elementos diagonais todos iguais a 1 e **u** é uma matriz escalonada. Se **a** for uma matriz quadrada $m \times m$ então $\mathbf{u} \in M(m \times m)$ será uma matriz triangular superior. Esta é a chamada *decomposição* **lu** da matriz **a**.

É importante observar os seguintes fatos a respeito da decomposição $\mathbf{a} = \mathbf{lu}$.

1º) O método gaussiano de eliminação fornece diretamente os elementos das matrizes **l** e **u**.

2º) Quando se dispõe da decomposição $\mathbf{a} = \mathbf{lu}$, a solução do sistema $\mathbf{ax} = \mathbf{b}$, com $\mathbf{b} \in M(m \times 1)$, se reduz a resolver o sistema $\mathbf{ly} = \mathbf{b}$ e, depois de obtida **y**, o sistema $\mathbf{ux} = \mathbf{y}$. O primeiro se resolve de cima para baixo e o segundo de baixo para cima, pois **l** é triangular inferior e **u** é escalonada.

3º) A maior vantagem computacional da decomposição **a**=**lu** ocorre quando se tem de resolver um grande número de equações $\mathbf{ax} = \mathbf{b}$, com a mesma matriz **a** e muitos vetores b. Dispondo da decomposição $\mathbf{a} = \mathbf{lu}$ não é preciso repetir muitas vezes o processo de eliminação gaussiana.

4º) A priori, não se sabe quais são (nem mesmo se vão ser necessárias) as transposições de linhas durante o processo de escalonamento de uma matriz **a**. Entretanto, depois de efetuado o processo, dispomos da relação de todas as transposições feitas. Efetuando, na mesma ordem em que foram feitas, todas essas transposições nas linhas da matriz identidade, obtemos uma matriz $\mathbf{p} \in M(m \times m)$, que se chama uma *matriz de permutação*. O produto **pa** corresponde a efetuar sobre a matriz **a**, antecipadamente, todas as transposições de linhas que seriam necessárias durante o escalonamento. Por-

tanto a matriz **pa** pode ser escalonada usando apenas operações elementares do tipo $L_i - \alpha L_j$. Assim, tem-se a decomposição **pa** $=$ **lu**.

5º) Uma condição suficiente (mas não necessária) para que uma matriz $\mathbf{a} \in M(m \times n)$ possa ser escalonada sem transposições de linhas, portanto admita uma decomposição $\mathbf{A} = \mathbf{lu}$, é que suas submatrizes principais sejam invertíveis. (Para todo número natural $r \leq \min\{m, n\}$, a *submatriz principal* de ordem r da matriz $\mathbf{a} = [a_{ij}]$ é a matriz $\mathbf{a}_r \in M(r \times r)$ formada pelos elementos a_{ij} com $1 \leq i \leq r$ e $1 \leq j \leq r$.)

Provemos este fato no caso de uma matriz $\mathbf{a} \in M(3 \times 4)$. O argumento no caso geral é o mesmo mas a notação se torna mais complicada e pode obscurecer a idéia, que é extremamente simples. Seja

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}.$$

Como a matriz principal $[a_{11}]$ é invertível, temos $a_{11} \neq 0$. Usando a_{11} como pivô, efetuamos as operações elementares

$$L_2 - \frac{a_{21}}{a_{11}} L_1 \quad \text{e} \quad L_3 - \frac{a_{31}}{a_{11}} L_1$$

sobre a matriz **a**, obtendo a nova matriz

$$\mathbf{a}^{(1)} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & b_{22} & b_{23} & b_{24} \\ 0 & b_{32} & b_{33} & b_{34} \end{bmatrix}.$$

A matriz

$$\begin{bmatrix} a_{11} & a_{12} \\ 0 & b_{22} \end{bmatrix}$$

resulta da matriz principal

$$\mathbf{a}_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

pela aplicação da operação elementar $L_2 - (a_{21}/a_{11})L_1$, logo é invertível. Como é triangular, isto obriga $b_{22} \neq 0$. Assim podemos tomar b_{22} como pivô. (Na notação geral, seria $a_{22}^{(1)}$ em vez de b_{22}.)

Aplicando a operação $L_2 - (b_{32}/b_{22})L_2$ à matriz $\mathbf{a}^{(1)}$ chegamos à matriz escalonada

$$\mathbf{a}^{(2)} = \mathbf{u} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & b_{22} & b_{23} & b_{24} \\ 0 & 0 & c_{33} & c_{34} \end{bmatrix}$$

O mesmo raciocínio se aplica em geral.

Exemplo 17.1. Considerando o escalonamento

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 12 & 12 \end{bmatrix} \underset{L_3-9L_1}{\overset{L_2-5L1}{\longrightarrow}} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & -8 & -15 & -24 \end{bmatrix} \overset{L_3-2L_2}{\longrightarrow}$$

$$\overset{L_3-2L_2}{\longrightarrow} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

obtemos a decomposição

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 12 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 9 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

que exprime a matriz do primeiro membro como um produto do tipo **lu**, de uma matriz triangular inferior com diagonal $(1, 1, 1)$ por uma matriz escalonada com pivôs $1, -4, 1$. (Todos $\neq 0$ pois a matriz dada tem posto máximo.)

Exemplo 17.2. Examinando, no Exemplo 9.4, as operações elementares efetuadas a partir da segunda, quando não ocorrem mais transposições de linhas, obtemos a seguinte decomposição do tipo $\mathbf{a} = \mathbf{lu}$:

$$\begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 3 & 4 & 2 & 0 \\ 4 & 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{3}{2} & \frac{5}{2} & 1 & 0 \\ 2 & 0 & \frac{4}{5} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -\frac{15}{2} & -\frac{15}{2} \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

17.G Matrizes Positivas: Cholesky versus lu

Quando a matriz $\mathbf{a} = [a_{ij}] \in M(m \times m)$ é positiva, afirmamos que todas as submatrizes principais $\mathbf{a}_r = [a_{ij}]$, $1 \le i, j \le r$, são também positivas, portanto invertíveis. Com efeito, dado o vetor não-nulo $v = (x_1 \ldots, x_r)$, tomamos $\overline{v} = (\overline{x}_1, \ldots, \overline{x}_m)$, com $\overline{x}_i = x_i$ para $i = 1, \ldots, r$ e $\overline{x}_i = 0$ para $i = r+1, \ldots, m$. Então $\overline{v}$ é um vetor não-nulo em $\mathbb{R}^m$, logo

$$\sum_{i,j=1}^{r} a_{ij} x_i x_j = \sum_{i,j=1}^{m} a_{ij} \overline{x}_i \overline{x}_j > 0.$$

Portanto o processo de escalonamento, quando aplicado a uma matriz positiva $\mathbf{a}$, não requer transposições de linhas. Assim, tem-se a decomposição $\mathbf{a} = \mathbf{lu}$.

Vemos que toda matriz positiva possui duas decomposições: Cholesky e $\mathbf{lu}$. É natural indagar qual a relação entre elas. Mostraremos que são muito próximas: qualquer delas se obtém a partir da outra de modo simples. Antes porém precisamos estabelecer a unicidade abaixo.

Se a matriz $\mathbf{a}$ é invertível então existe uma única maneira de escrevê-la como um produto $\mathbf{a} = \mathbf{lu}$, onde $\mathbf{l}$ é triangular inferior, com os elementos da diagonal todos iguais a 1, e $\mathbf{u}$ é triangular superior.

Com efeito, sendo $\mathbf{a}$ invertível, a igualdade $\mathbf{a} = \mathbf{lu}$ obriga $\mathbf{l}$ e $\mathbf{u}$ a serem invertíveis. Portanto

$$\mathbf{a} = \mathbf{l}_1\mathbf{u}_1 = \mathbf{l}_2\mathbf{u}_2 \Rightarrow \mathbf{l}_2^{-1}\mathbf{l}_1 = \mathbf{u}_2\mathbf{u}_1^{-1}.$$

Ora, o primeiro membro da última igualdade é uma matriz triangular inferior e o segundo membro é triangular superior. Logo ambas são matrizes diagonais. Como os elementos diagonais de $\mathbf{l}$ são todos iguais a 1, segue-se que $\mathbf{l}_1 = \mathbf{l}_2 = \mathbf{I}_m$, portanto $\mathbf{u}_1 = \mathbf{u}_2$ e fica provada a unicidade da decomposição $\mathbf{a} = \mathbf{lu}$ no caso de $\mathbf{a}$ ser invertível.

Voltemos à matriz positiva $\mathbf{a} \in M(m \times m)$, que admite as decomposições $\mathbf{a} = \mathbf{t}^T.\mathbf{t} = \mathbf{l}.\mathbf{u}$, onde $\mathbf{l}$ é triangular inferior com 1's na diagonal, $\mathbf{t}$ e $\mathbf{u}$ são triangulares superiores e todos os elementos da diagonal de $\mathbf{t}$ são positivos. Se indicarmos com $\mathbf{d} \in M(m \times m)$ a matriz diagonal com $d_{ii} = t_{ii}$, então $\mathbf{a} = (\mathbf{t}^T\mathbf{d}^{-1})(\mathbf{dt})$, onde $\mathbf{t}^T\mathbf{d}^{-1}$ é triangular inferior com 1's na diagonal e $\mathbf{dt}$ é triangular superior. Segue-se da unicidade acima provada que $\mathbf{l} = \mathbf{t}^T\mathbf{d}^{-1}$ e $\mathbf{u} = \mathbf{dt}$, mostrando as-

sim como obter a decomposição $\mathbf{a} = \mathbf{lu}$ a partir da decomposição de Cholesky $\mathbf{a} = \mathbf{t}^\top.\mathbf{t}$.

A expressão $\mathbf{u} = \mathbf{dt}$ mostra, em particular que se a matriz $\mathbf{a}$ é positiva então, em sua decomposição $\mathbf{a} = \mathbf{lu}$, os elementos u_{ii} da diagonal de $\mathbf{u}$ (isto é, os *pivôs* da eliminação gaussiana) são todos positivos. A unicidade da decomposição $\mathbf{lu}$ implica que esses pivôs são univocamente determinados a partir da matriz $\mathbf{a}$.

Mostremos agora como se pode obter a decomposição de Cholesky de uma matriz positiva a partir de sua decomposição $\mathbf{a} = \mathbf{lu}$.

Como $u_{ii} > 0$ para $i = 1, \ldots, m$, podemos formar uma matriz diagonal $\mathbf{d} = [d_{ij}]$ pondo $d_{ij} = 0$ se $i \neq j$ e $d_{ii} = \sqrt{u_{ii}}$. Então escrevemos $\mathbf{t} = \mathbf{d}^{-1}\mathbf{u}$. Evidentemente, $\mathbf{a} = (\mathbf{ld})(\mathbf{d}^{-1}\mathbf{u})$. Portanto teremos a decomposição de Cholesky $\mathbf{a} = \mathbf{t}^\top.\mathbf{t}$ desde que provemos que $\mathbf{t}^\top = \mathbf{ld}$, ou seja, que $\mathbf{u}^\top.\mathbf{d}^{-1} = \mathbf{ld}$.

Ora, se tomarmos a transposta de ambos os membros da igualdade $\mathbf{a} = \mathbf{lu}$ obteremos $\mathbf{a} = \mathbf{a}^\top = \mathbf{u}^\top\mathbf{l}^\top$, logo podemos escrever $\mathbf{a} = \mathbf{u}^\top.\mathbf{d}^{-2}.\mathbf{d}^2.\mathbf{l}^\top$. Mas $\mathbf{u}^\top.\mathbf{d}^{-2}$ é triangular inferior com 1's na diagonal. (Na realidade, $\mathbf{d}$ foi definida com esse propósito.) E $\mathbf{d}^2\mathbf{l}^\top$ é triangular superior. Pela unicidade da decomposição $\mathbf{a} = \mathtt{boldl}\mathbf{u}$, vem $\mathbf{u}^\top.\mathbf{d}^{-2} = \mathbf{l}$ (e $\mathbf{d}^2.\mathbf{l}^\top = \mathbf{u}$). Daí resulta $\mathbf{u}^\top.\mathbf{d}^{-1} = \mathbf{ld}$, como queríamos provar.

Em suma: se $\mathbf{a} = \mathbf{lu}$ é a decomposição de uma matriz positiva, sua decomposição de Cholesky é $\mathbf{a} = (\mathbf{ld})(\mathbf{d}^{-1}\mathbf{u})$, onde $\mathbf{d}$ é a matriz diagonal formada pelas raízes quadradas dos elementos (necesssariamente positivos) da diagonal de $\mathbf{u}$. Isto nos dá um método prático para obter a decomposição de Cholesky, como veremos no Exemplo 17.3.

Observação O argumento acima prova, na realidade, que se $\mathbf{a}$ é uma matriz simétrica cujos pivôs são todos positivos então existe uma decomposição $\mathbf{a} = \mathbf{t}^\top.\mathbf{t}$, logo $\mathbf{a}$ é uma matriz positiva. (Observe que, os pivôs sendo positivos, a matriz $\mathbf{a}$ é invertível.)

Exemplo 17.3. Como exemplo de matriz positiva, tomemos a matriz de Gram dos vetores $u = (1,2,1)$, $v = (2,1,2)$, $w = (1,1,2)$. Escalonando-se, obtemos

$$\begin{bmatrix} 6 & 6 & 5 \\ 6 & 9 & 7 \\ 5 & 7 & 6 \end{bmatrix} \xrightarrow[L_3-\frac{5}{6}L_1]{L_2-L_1} \begin{bmatrix} 6 & 6 & 5 \\ 0 & 3 & 2 \\ 0 & 2 & 11/6 \end{bmatrix} \xrightarrow{L_3-\frac{2}{3}L_2} \begin{bmatrix} 6 & 6 & 5 \\ 0 & 3 & 2 \\ 0 & 0 & 1/2 \end{bmatrix}$$

Isto nos fornece a decomposição **lu**:

$$\begin{bmatrix} 6 & 6 & 5 \\ 6 & 9 & 7 \\ 5 & 7 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{5}{6} & \frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 6 & 6 & 5 \\ 0 & 3 & 2 \\ 0 & 0 & 1/2 \end{bmatrix}.$$

Para obter a decomposição de Cholesky, tomamos a matriz diagonal

$$\mathbf{d} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{2}/2 \end{bmatrix}$$

e calculamos

$$\mathbf{t}^\mathsf{T} = \mathbf{ld} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{5}{6} & \frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ \sqrt{6} & \sqrt{3} & 0 \\ \frac{5}{\sqrt{6}} & \frac{2}{\sqrt{3}} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

Então a decomposição de Cholesky da matriz dada é

$$\begin{bmatrix} 6 & 6 & 5 \\ 6 & 9 & 7 \\ 5 & 7 & 6 \end{bmatrix} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ \sqrt{6} & \sqrt{3} & 0 \\ 5/\sqrt{6} & 2/\sqrt{3} & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} \sqrt{6} & \sqrt{6} & 5/\sqrt{6} \\ 0 & \sqrt{3} & 2/\sqrt{3} \\ 0 & 0 & \sqrt{2}/2 \end{bmatrix}.$$

Exercícios

17.1. Prove que os vetores $v_1, \dots, v_k \in E$ geram um subespaço vetorial de dimensão r se, e somente se, a matriz de Gram $\mathbf{g}(v_1, \dots, v_k)$ tem posto r.

17.2. Se $\dim E \le \dim F$, prove que existe uma transformação linear ortogonal $A \colon E \to F$ tal que $Av_1 = w_1, \dots, Av_k = w_k$ se, e somente se, as matrizes de Gram $\mathbf{g}(v_1, \dots, v_k)$ e $\mathbf{g}(w_1, \dots, w_k)$ são iguais.

17.3. Sejam $\mathbf{a} \in M(n \times n)$ uma matriz positiva e $\mathbf{b} \in M(n \times m)$ uma matriz de posto n. Ache vetores $v_1, \dots, v_m \in \mathbb{R}^n$ tais que $\mathbf{g}(v_1, \dots, v_m) = \mathbf{b}^\mathsf{T}\mathbf{ab}$.

17.4. Prove que $\mathbf{a}^\mathsf{T}\mathbf{a}$ é a matriz de Gram dos vetores-coluna de $\mathbf{a}$.

17.5. Um operador $T: E \to E$ chama-se *triangularizável* quando o espaço vetorial E possui uma base, relativamente à qual a matriz de T é triangular. Prove:

(a) T é triangularizável se, e somente se, existe uma cadeia ascendente $\{0\} = F_o \subset F_1 \subset \cdots \subset F_n = E$, de subespaços invariantes por T, com $\dim F_i = i \ (i = 0, \ldots, n)$.

(b) T é triangularizável se, e somente se, existe uma cadeia descendente $E = G_n \supset \cdots \supset G_1 \supset G_o = \{0\}$ de subespaços invariantes por T tais que $\dim G_i = i$ para $i = 0, 1, \ldots, n$.

(c) Se existe uma base de E na qual a matriz de T é triangular superior, existe também uma base de E na qual a matriz de T é triangular inferior.

(d) T é triangularizável se, e somente se, T^* é triangularizável.

17.6. Seja $\mathbf{t} = [t_{ij}] \in M(n \times n)$ uma matriz triangular. Prove que os elementos t_{ii} da diagonal de $\mathbf{t}$ são autovalores do operador $T: \mathbb{R}^n \to \mathbb{R}^n$ cuja matriz na base canônica é $\mathbf{t}$. Conclua que $\mathbf{t}$ é diagonalizável quando sua diagonal não possuir elementos repetidos.

17.7. Seja $A: \mathbb{R}^3 \to \mathbb{R}^3$ o operador definido por $A(x, y, z) = (x + 2y + 3z, y + 2z, 3z)$. Obtenha dois autovetores L.I. para A e prove que qualquer outro autovetor de A é múltiplo de um desses dois. Conclua que A não é diagonalizável, embora sua matriz seja triangular.

17.8. Considere o operador $B: \mathbb{R}^3 \to \mathbb{R}^3$, dado por $B(x, y, z) = (x + 2z, y+3z, 4z)$, cuja matriz é triangular. Ache uma base de $\mathbb{R}^3$ formada por autovetores de B.

17.9. Dada a matriz positiva

$$\mathbf{a} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix},$$

determine

$$\mathbf{t} = \begin{bmatrix} x & y \\ 0 & z \end{bmatrix}$$

de modo que se tenha a decomposição de Cholesky $\mathbf{a} = \mathbf{t}^T.\mathbf{t}$.

17.10. Use o método do exercício anterior, de modo a obter a decomposição de Cholesky da matriz positiva

$$\mathbf{a} = \begin{bmatrix} 3 & 4 & 6 \\ 4 & 6 & 9 \\ 6 & 9 & 14 \end{bmatrix}.$$

17.11. Aplicando o processo de Gram-Schmidt às colunas da matriz

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 3 & 4 \\ 4 & -1 & 3 \end{bmatrix},$$

obtenha a decomposição $\mathbf{a} = \mathbf{qr}$, onde $\mathbf{q}$ é ortogonal e $\mathbf{r}$ é triangular superior, com elementos positivos na diagonal.

17.12. Mostre como obter, a partir dos teoremas demonstrados nas seções anteriores, cada uma das matrizes cuja existência é assegurada nos ítens 1), 2), 3) e 4) de **17.E** (Por exemplo, no item 1), se $A\colon \mathbb{R}^n \to \mathbb{R}^n$ é o operador cuja matriz na base canônica $\mathcal{E} \subset \mathbb{R}^n$ é $\mathbf{a}$ então $\mathbf{p}$ é a matriz cujas colunas são os vetores de uma base ortonormal $\mathcal{U} = \{u_1, \ldots, u_n\} \subset \mathbb{R}^n$, formada por autovetores de A e $\mathbf{d}$ é a matriz diagonal formada pelos autovalores correspondentes.)

17.13. Dada a matriz

$$\mathbf{a} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix},$$

obtenha sua decomposição $\mathbf{a} = \mathbf{pdq}$ a valores singulares.

17.14. Assinale V(erdadeiro) ou F(also)

( ) Toda matriz quadrada escalonada é triangular superior.

( ) Toda matriz (quadrada) triangular superior é escalonada.

( ) As duas afirmações anteriores são verdadeiras quando se trata de matrizes invertíveis.

( ) Se a matriz $\mathbf{a}$ admite uma decomposição do tipo $\mathbf{a} = \mathbf{lu}$ então todas as submatrizes principais de $\mathbf{a}$ são invertíveis.

17.15. Prove que a matriz

$$\mathbf{a} = \begin{bmatrix} 0 & a \\ b & c \end{bmatrix}$$

admite uma decomposição do tipo $\mathbf{a} = \mathbf{lu}$ se, e somente se, $b = 0$.

17.16. Obtenha a decomposição **lu** das matrizes

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 3 & 4 \\ 4 & -1 & 3 \end{bmatrix} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} 2 & 1 & 0 & 4 \\ 4 & 5 & 1 & 7 \\ 2 & -8 & -1 & 12 \end{bmatrix}.$$

17.17. Seja $\mathbf{a} \in M(m \times n)$ uma matriz de posto máximo que admite uma decomposição do tipo $\mathbf{a} = \mathbf{lu}'$, onde $\mathbf{l} \in M(m \times m)$ é triangular inferior, com elementos da diagonal todos iguais a 1, e $\mathbf{u}' \in M(m \times n)$ é escalonada. Prove que existem uma matriz diagonal invertível $\mathbf{d} \in M(m \times m)$ e uma matriz escalonada $\mathbf{u} \in M(m \times n)$, com pivôs todos igual a 1, tais que $\mathbf{a} = \mathbf{ldu}$.

17.18. Obtenha a decomposição **lu** da matriz do Exemplo 9.3.

17.19. A partir da decomposição **lu**, obtenha a decomposição de Cholesky das matrizes positivas $\mathbf{a}$ = matriz de Gram dos vetores $u = (1, 2, 1)$, $v = (1, 1, 2)$, $w = (2, 1, 3)$ e $\mathbf{b}$ = produto da matriz

$$\mathbf{m} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 2 & 1 & 2 & 2 \\ 2 & 2 & 1 & 3 \end{bmatrix}$$

por sua transposta.

17.20. Foi visto no texto que se $\mathbf{a} = \mathbf{lu}$ é a decomposição da matriz positiva $\mathbf{a}$ e $\mathbf{d}$ é a matriz diagonal formada pelas raízes quadradas $d_{ii} = \sqrt{u_{ii}}$ dos pivôs então $\mathbf{a} = (\mathbf{ld}) \cdot (\mathbf{d}^{-1}\mathbf{u})$ é a decomposição de Cholesky de $\mathbf{a}$. Modifique ligeiramente este argumento para provar que existe uma decomposição de Cholesky $\mathbf{a} = \mathbf{t}^T.\mathbf{t}$ para qualquer matriz não-negativa $\mathbf{a}$. (Mas agora não há unicidade.)

18

Formas Quadráticas

Uma forma quadrática num espaço vetorial E é uma função que, em termos das coordenadas de um vetor relativamente a uma base de E, se exprime como um polinômio homogêneo do segundo grau. As formas quadráticas ocorrem com grande destaque em problemas de otimização (máximos e mínimos), no estudo das superfícies quádricas, na Geometria Diferencial, na Mecânica, etc. Em todas essas situações, é relevante o conhecimento do sinal (positivo ou negativo) que a forma pode assumir ou, mais precisamente, dos seus autovalores. Nesta seção, é feito um estudo conciso, porém abrangente, dos principais pontos básicos referentes a essas funções e de suas relações com os operadores lineares, finalizando com o método de Lagrange para diagonalização e a classificação das superfícies quádricas.

Sejam E, F espaços vetoriais. Uma *forma bilinear* $b\colon E \times F \to \mathbb{R}$ é uma função $b(u, v)$, linear em cada uma das duas variáveis $u \in E$, $v \in F$. Mais precisamente, para quaisquer $u, u' \in E$, $v, v' \in F$ e $\alpha \in \mathbb{R}$ devem valer:

$$b(u + u', v) = b(u, v) + b(u', v); \qquad b(\alpha u, v) = \alpha b(u, v)$$
$$b(u, v + v') = b(u, v) + b(u, v'); \qquad b(u, \alpha v) = \alpha b(u, v).$$

As operações evidentes de soma e produto por um número fazem do conjunto $\mathcal{B}(E \times F)$ das formas bilineares $b\colon E \times F \to \mathbb{R}$ um espaço vetorial.

Tomando bases $\mathcal{U} = \{u_1, \ldots, u_m\} \subset E$ e $\mathcal{V} = \{v_1, \ldots, v_n\} \subset F$, os números $b_{ij} = b(u_i, v_j)$ definem uma matriz $\mathbf{b} = [b_{ij}] \in M(m \times n)$, chamada a *matriz da forma bilinear* b *relativamente às bases* $\mathcal{U}$, $\mathcal{V}$.

Conhecidos os valores $b_{ij} = b(u_i, v_j)$, os quais podem ser tomados arbitrariamente, a forma bilinear $b\colon E \times F \to \mathbb{R}$ fica inteiramente determinada pois, para $u = \Sigma x_i u_i \in E$ e $v = \Sigma y_j v_j \in F$ quaisquer, tem-se

$$b(u, v) = \sum_{i,j} x_i y_j \, b(u_i, v_j) = \sum_{i,j} b_{ij} x_i y_j, \quad 1 \leq i \leq m, \ 1 \leq j \leq n.$$

A correspondência que associa a cada forma bilinear $b\colon E \times F \to \mathbb{R}$ sua matriz $\mathbf{b} = [b_{ij}]$ relativamente às bases $\mathcal{U} \subset E$, $\mathcal{V} \subset F$ (supostas fixadas) é um isomorfismo entre os espaços vetoriais $\mathcal{B}(E \times F)$ e $M(m \times n)$. Segue-se que $\mathcal{B}(E \times F)$ tem dimensão mn.

Dadas novas bases $\mathcal{U}' = \{u'_1, \ldots, u'_m\} \subset E$ e $\mathcal{V}' = \{v'_1, \ldots, v'_n\} \subset F$ sejam

$$u'_j = \sum_{i=1}^{m} p_{ij} u_i, \quad v'_j = \sum_{i=1}^{n} q_{ij} v_i$$

e $b'_{ij} = b(u'_i, v'_j)$. A matriz da forma bilinear b relativamente às bases $\mathcal{U}'$ e $\mathcal{V}'$ é $\mathbf{b}' = [b'_{ij}] \in M(m \times n)$. O teorema abaixo exprime $\mathbf{b}'$ em função de $\mathbf{b}$ e das matrizes de passagem $\mathbf{p} = [p_{ij}] \in M(m \times m)$, $\mathbf{q} = [q_{ij}] \in M(n \times n)$.

Teorema 18.1. *As matrizes* $\mathbf{b}$ *e* $\mathbf{b}'$, *da forma bilinear* b *nas bases* $\mathcal{U}$, $\mathcal{V}$ *e* $\mathcal{U}'$, $\mathcal{V}'$ *respectivamente, se relacionam pela igualdade* $\mathbf{b}' = \mathbf{p}^T \mathbf{b} \mathbf{q}$, *onde* $\mathbf{p}$ *é a matriz de passagem de* $\mathcal{U}$ *para* $\mathcal{U}'$ *e* $\mathbf{q}$ *é a matriz de passagem de* $\mathcal{V}$ *para* $\mathcal{V}'$.

Demonstração: Para todo $i = 1, \ldots, m$ e todo $j = 1, \ldots, n$, temos

$$\begin{aligned} b'_{ij} = b(u'_i, v'_j) &= b\left(\sum_{r=1}^{m} p_{ri} u_r, \sum_{s=1}^{n} q_{sj} v_s\right) = \sum_{r,s} p_{ri} q_{sj} \, b(u_r, v_s) \\ &= \sum_{r,s} p_{ri} q_{sj} b_{rs} = \sum_{r,s} p_{ri} b_{rs} q_{sj} = (\mathbf{p}^T \mathbf{b} \mathbf{q})_{ij} \,, \end{aligned}$$

logo $\mathbf{b}' = \mathbf{p}^T \mathbf{b} \mathbf{q}$. □

Embora transformações lineares e formas bilineares sejam ambas representáveis por matrizes (quando são fixadas bases), vemos que, ao mudarem essas bases, a matriz de uma forma bilinear se altera de maneira diferente daquela de uma transformação linear.

Quando $E = F$, nos referiremos sempre (salvo aviso em contrário) à matriz de uma forma bilinear $b\colon E \times E \to \mathbb{R}$ relativamente a uma única base $\mathcal{U} = \{u_1, \ldots, u_m\} \subset E$. Essa matriz é $\mathbf{b} = [b_{ij}] \in M(m \times m)$, onde $b_{ij} = b(u_i, u_j)$.

Se $\mathbf{b}'$ é a matriz da mesma forma b em relação a outra base $\mathcal{U}' \subset E$ então $\mathbf{b}' = \mathbf{p}^T \mathbf{b} \mathbf{p}$, onde $\mathbf{p}$ é a matriz de passagem da base $\mathcal{U}$ para a base $\mathcal{U}'$. A correspondência $b \mapsto \mathbf{b}$ define um isomorfismo entre $\mathcal{B}(E \times E)$ e $M(m \times m)$, estabelecido com ajuda da base $\mathcal{U}$. Dados $u = \Sigma x_i u_i$ e $v = \Sigma y_i u_i$ em E, se a matriz de b na base $\mathcal{U}$ é $\mathbf{b} = [b_{ij}]$, tem-se

$$b(u, v) = \sum_{i,j=1}^{m} b_{ij} x_i y_j \,.$$

Assim, $b(u, v)$ é um polinômio homogêneo do 2º grau em relação às coordenadas de u e v.

Uma forma bilinear $b\colon E \times E \to \mathbb{R}$ chama-se *simétrica* quando $b(u, v) = b(v, u)$ para quaisquer $u, v \in E$. Para que b seja simétrica é suficiente que sua matriz em relação a uma base $\mathcal{U} \subset E$ seja simétrica e é necessário que sua matriz em relação a qualquer base de E seja simétrica. Com efeito, se $b_{ij} = b_{ji}$ então

$$\begin{aligned} b(v, u) = \sum_{i,j} b_{ij} y_i x_j &= \sum_{i,j} b_{ij} x_j y_i \\ &= \sum_{i,j} b_{ji} x_j y_i = \sum_{\alpha,\beta} b_{\alpha\beta} x_\alpha y_\beta = b(u, v). \end{aligned}$$

Observação. O quarto sinal de igualdade na seqüência acima faz uso de uma manipulação que pode parecer desonesta mas é perfeitamente correta. O princípio é o seguinte: num somatório, o nome do "índice de somação", ou "índice mudo", não tem a menor importância:

$$\sum_{i=1}^{m} z_i = \sum_{\alpha=1}^{m} z_\alpha = \sum_{r=1}^{m} z_r = z_1 + \cdots + z_m.$$

O que importa é que ao substituir o índice mudo i por β (bem como j por α) essa substituição se faça em todos os lugares onde i ocorra (assim como j) e somente nesses lugares.

Analogamente, uma forma bilinear $b\colon E \times E \to \mathbb{R}$ chama-se *anti-simétrica* quando $b(u,v) = -b(v,u)$ para $u,v \in E$ quaisquer. Para que b seja anti-simétrica é suficiente que se tenha $b(u_i,u_j) = -b(u_j,u_i)$ ou seja, $b_{ij} = -b_{ji}$ numa base $\{u_1,\dots,u_m\} \subset E$.

Uma forma bilinear b, ao mesmo tempo simétrica e anti-simétrica, deve ser nula. Este fato, juntamente com a igualdade

$$b(u,v) = \frac{1}{2}[b(u,v) + b(v,u)] + \frac{1}{2}[b(u,v) - b(v,u)],$$

mostra que o espaço $\mathcal{B}(E \times E)$ das formas bilineares $b\colon E \times E \to \mathbb{R}$ é a soma direta de dois subespaços, um deles formado pelas formas simétricas e o outro pelas formas anti-simétricas.

Exemplo 18.1. Dados os funcionais lineares $f\colon E \to \mathbb{R}$, $g\colon F \to \mathbb{R}$, a função $b\colon E \times F \to \mathbb{R}$, definida por $b(u,v) = f(u)g(v)$, é uma forma bilinear, chamada o *produto tensorial* de f e g. Se E e F são dotados de produto interno, fixados $u_o \in E$ e $v_o \in F$, a função $b\colon E \times F \to \mathbb{R}$, onde $b(u,v) = \langle u,u_o\rangle \cdot \langle v,v_o\rangle$, é uma forma bilinear. No caso particular em que $E = F$, dados $f,g\colon E \to \mathbb{R}$, funcionais lineares, então as igualdades

$$(f \bullet g)(u,v) = f(u)g(v) + f(v)g(u)$$

e

$$(f \wedge g)(u,v) = f(u)g(v) - f(v)g(u),$$

definem formas bilineares $f \bullet g, f \wedge g\colon E \times E \to \mathbb{R}$, a primeira simétrica e a segunda anti-simétrica.

Teorema 18.2. *Seja E um espaço vetorial de dimensão finita provido de produto interno. Para cada forma bilinear $b\colon E \times E \to \mathbb{R}$ existe um único operador linear $B\colon E \to E$ tal que*

$$\langle u, Bv\rangle = b(u,v) \quad \textit{para } u,v \in E \textit{ quaisquer}.$$

A correspondência $b \mapsto B$, assim definida, é um isomorfismo entre os espaços vetoriais $\mathcal{B}(E \times E)$ e $\mathcal{L}(E)$. A forma bilinear b é simétrica (respect. anti-simétrica) se, e somente se, o operador B é auto-adjunto (respect. anti-simétrico).

Demonstração: Para cada $v_o \in E$, a função $f\colon E \to \mathbb{R}$, definida por $f(u) = b(u, v_o)$, é um funcional linear. Pelo Teorema 11.1, existe um único vetor em E, que indicaremos com Bv_o, tal que $f(u) = \langle u, Bv_o \rangle$, ou seja $\langle u, Bv_o \rangle = b(u, v_o)$ para todo $u \in E$. Como $v_o \in E$ é arbitrário, acabamos de mostrar que existe uma única função $B\colon E \to E$ tal que $\langle u, Bv \rangle = b(u, v)$ para quaisquer $u, v \in E$. Dados $v, v' \in E$, tem-se

$$\begin{aligned}\big\langle u, B(v+v')\big\rangle &= b(u, v+v') \\ &= b(u,v) + b(u,v') \\ &= \langle u, Bv\rangle + \big\langle u, Bv'\big\rangle \\ &= \big\langle u, Bv + Bv'\big\rangle\end{aligned}$$

para todo $u \in E$, logo $B(v+v') = Bv+Bv'$. De modo análogo se verifica que $B(\alpha v) = \alpha(Bv)$ para $\alpha \in \mathbb{R}$ e $v \in E$ quaisquer, portanto $B\colon E \to E$ é linear. Em relação a uma base ortonormal $\mathcal{U} = \{u_1, \dots, u_m\} \subset E$, o ij-ésimo elemento da matriz de B é $\langle e_i, Be_j \rangle = b(e_i, e_j) =$ ij-ésimo da matriz de b. Assim, a forma bilinear b e o operador B que a ela corresponde têm a mesma matriz na base $\mathcal{U}$. Daí decorre que a correspondência $b \mapsto B$ é um isomorfismo entre $\mathcal{B}(E \times E)$ e $\mathcal{L}(E)$ e que B é auto-adjunto (respect. anti-simétrico) se,e somente se, b é simétrica (respect. anti-simétrica). $\square$

Uma função $\varphi\colon E \to \mathbb{R}$ chama-se uma *forma quadrática* quando existe uma forma bilinear $b\colon E \times E \to \mathbb{R}$ tal que $\varphi(v) = b(v, v)$ para todo $v \in E$.

Se, em vez da forma bilinear b tomarmos a forma bilinear simétrica

$$b'(u, v) = \frac{1}{2}[b(u, v) + b(v, u)],$$

teremos ainda

$$\varphi(v) = b(v, v) = \frac{1}{2}[b(v, v) + b(v, v)] = b'(v, v).$$

Portanto, não há perda de generalidade em se exigir que a forma quadrática $\varphi(v) = b(v, v)$ provenha de uma forma bilinear simétrica b. Isto é o que faremos doravante. Com uma vantagem: se b é simétrica, todos os seus valores $b(u, v)$ podem ser determinados a partir dos valores $b(v, v) = \varphi(v)$ da forma quadrática φ. Com efeito, tem-se a seguinte fórmula de *polarização*:

$$b(u, v) = \frac{1}{2}[b(u+v, u+v) - b(u, u) - b(v, v)].$$

Se $\mathbf{b} = [b_{ij}]$ é a matriz da forma bilinear b na base $\mathcal{U} = \{u_1, \ldots, u_m\} \subset E$ então, para $v = \Sigma x_i u_i$, tem-se

$$\varphi(v) = \sum_{i,j=1}^{m} b_{ij} x_i x_j \, .$$

Note que, sendo $\mathbf{b}$ uma matriz simétrica (como sempre suporemos quando tratarmos de formas quadráticas), cada parcela com $i \neq j$ aparece duas vezes na soma acima: uma vez como $b_{ij}x_ix_j$ e outra vez como $b_{ji}x_jx_i$, o que é o mesmo.

A *matriz da forma quadrática* φ na base $\mathcal{U}$ é, por definição, a matriz $\mathbf{b}$, nesta mesma base, da forma bilinear b tal que $\varphi(v) = b(v, v)$. Se a matriz de passagem $\mathbf{p}$ levar a base $\mathcal{U}$ na base $\mathcal{U}'$, a matriz $\mathbf{b}'$ da forma quadrática φ na base $\mathcal{U}'$ será dada por $\mathbf{b}' = \mathbf{p}^T\mathbf{b}\mathbf{p}$.

Note que se E possuir produto interno e as bases $\mathcal{U}, \mathcal{U}'$ forem ambas ortonormais, a matriz $\mathbf{p}$ será ortogonal, logo $\mathbf{p}^T = \mathbf{p}^{-1}$. Então $\mathbf{b}' = \mathbf{p}^{-1}\mathbf{b}\mathbf{p}$. Isto confirma o que foi visto no Teorema 18.1: relativamente a bases ortonormais a matriz da forma bilinear b coincide com a do operador B, que lhe corresponde de maneira intrínseca.

Teorema 18.3. *Seja E um espaço vetorial de dimensão finita munido de produto interno. Dada uma forma bilinear simétrica $b\colon E \times E \to \mathbb{R}$, existe uma base ortonormal $\mathcal{U} = \{u_1, \ldots, u_m\} \subset E$ tal que $b(u_i, u_j) = 0$ se $i \neq j$.*

Demonstração: Pelo Teorema 18.2, existe um operador auto-adjunto $B\colon E \to E$ tal que $b(u, v) = \langle u, Bv \rangle$ para quaisquer $u, v \in E$. O Teorema Espectral assegura a existência de uma base ortonormal $\mathcal{U} = \{u_1, \ldots, u_m\} \subset E$ tal que $Bu_i = \lambda_i u_i$ $(i = 1, \ldots, m)$. Então $i \neq j \Rightarrow b(u_i, u_j) = \langle u_i, Bu_j \rangle = \langle u_i, \lambda_j u_j \rangle = \lambda_j \langle u_i, u_j \rangle = 0$. □

O teorema acima é a versão, para formas bilineares, do Teorema Espectral. Por sua vez a versão matricial do Teorema 18.3 é a seguinte: para toda matriz simétrica $\mathbf{b} = [b_{ij}] \in M(m \times m)$, existe uma matriz ortogonal $\mathbf{p} \in M(m \times m)$ tal que $\mathbf{p}^T\mathbf{b}\mathbf{p} = \mathbf{p}^{-1}\mathbf{b}\mathbf{p}$ é uma matriz diagonal. A diagonal de $\mathbf{d} = \mathbf{p}^T\mathbf{b}\mathbf{p}$ é formada pelos autovalores de $\mathbf{b}$, os quais se dizem também autovalores da forma bilinear b e da forma quadrática $\varphi(v) = b(v, v)$.

Em termos das coordenadas dos vetores $u = \Sigma x_i u_i$ e $v = \Sigma y_j u_j$ relativamente à base $\mathcal{U}$ do Teorema 18.3 a forma bilinear b se exprime como

$$b(u, v) = \Sigma \lambda_i x_i y_i \,.$$

Em particular, a forma quadrática $\varphi \colon E \to \mathbb{R}$, $\varphi(u) = b(u, u)$, para $v = \Sigma y_i u_i$ (expressão relativa à base $\mathcal{U}$) é dada por uma combinação linear de quadrados:

$$\varphi(v) = \Sigma \lambda_i y_i^2 = \lambda_1 y_1^2 + \cdots + \lambda_m y_m^2 \,.$$

É costume numerar os autovalores λ_i em ordem crescente: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$.

A forma quadrática $\varphi \colon E \to \mathbb{R}$ diz-se *não-negativa* quando $\varphi(v) \geq 0$ para todo $v \in E$, *positiva* quando $\varphi(v) > 0$ para todo $v \neq 0$ em E, *não-positiva* quando $\varphi(v) \leq 0$ para todo $v \in E$, *negativa* quando $\varphi(v) < 0$ para todo $v \neq 0$ em E e *indefinida* quando existem $u, v \in E$ tais que $\varphi(u) > 0$ e $\varphi(v) < 0$.

A forma quadrática $\varphi \colon E \to \mathbb{R}$ é não-negativa, positiva, não-positiva, negativa ou indefinida, respectivamente, conforme o operador auto-adjunto $B \colon E \to E$, tal que $\varphi(v) = \langle v, Bv \rangle$, tenha uma dessas propriedades, ou seja, conforme os autovalores $\lambda_1, \ldots, \lambda_m$ sejam todos ≥ 0, todos > 0, todos ≤ 0, todos < 0 ou $\lambda_1 < 0 < \lambda_m$ respectivamente.

Se $\lambda_1 \leq \cdots \leq \lambda_m$ são os autovalores da forma quadrática φ então, para todo vetor unitário $u \in E$ tem-se $\lambda_1 \leq \varphi(u) \leq \lambda_m$. Com efeito, relativamente à base $\mathcal{U}$ do Teorema 18.3, se $u = \Sigma x_i u_i$ então $\Sigma x_i^2 = 1$, portanto:

$$\lambda_1 = \sum_i \lambda_1 x_i^2 \leq \sum_i \lambda_i x_i^2 = \varphi(u) \leq \sum_i \lambda_m x_i^2 = \lambda_m \,.$$

Além disso, $\varphi(u_1) = \lambda_1$ e $\varphi(u_m) = \lambda_m$. Portanto, o menor autovalor λ_1 e o maior autovalor λ_m são também o valor mínimo e o valor máximo assumidos pela forma quadrática φ entre os vetores unitários de E.

Em Análise, quando se estudam os pontos críticos de uma função diferenciável, um papel crucial é desempenhado por uma forma quadrática, chamada forma Hessiana. Os pontos críticos são classificados de acordo com o número de direções independentes ao longo das

quais a função tem um máximo relativo e esse número coincide com o índice da forma Hessiana, como definiremos a seguir.

O *índice* de uma forma quadrática $\varphi \colon E \to \mathbb{R}$ é a maior dimensão de um subespaço vetorial de E restrita ao qual a forma é negativa. Quando $\varphi \geq 0$, diremos que o índice de φ é zero.

Portanto, o índice da forma quadrática $\varphi \colon E \to \mathbb{R}$ é igual ao número $i \neq 0$ quando: 1) Existe um subespaço vetorial $F \subset E$, com $\dim F = i$, tal que $\varphi(v) < 0$ para todo vetor não-nulo $v \in F$; 2) Se $G \subset E$ é um subespaço vetorial de dimensão maior do que i, existe algum vetor não-nulo $w \in G$, tal que $\varphi(w) \geq 0$.

Se $\dim E = m$ então toda forma quadrática negativa $\varphi \colon E \to \mathbb{R}$ tem índice m.

Assim definido, o índice de uma forma quadrática é uma noção intrínseca, independente de escolhas ou construções arbitrárias.

Outro invariante numérico associado a uma forma quadrática $\varphi \colon E \to \mathbb{R}$ é o seu *posto*. Seja $b \colon E \times E \to \mathbb{R}$ a (única) forma bilinear simétrica tal que $\varphi(v) = b(v, v)$ para todo $v \in E$. Escolhendo uma base $\mathcal{V} = \{v_1, \ldots, v_m\} \subset E$ e pondo $b_{ij} = b(v_i, v_j)$, $i, j = 1, \ldots, m$, sabemos que, para todo vetor

$$v = \sum x_i v_i, \quad \text{tem-se} \quad \varphi(v) = \sum_{i,j} b_{ij} x_i x_j.$$

Por definição, o *posto* de φ é o posto da matriz $\mathbf{b} = [b_{ij}] \in M(m \times m)$.

Para provar que o posto de φ, definido desta maneira, não depende da base escolhida, observemos que se tomarmos outra base $\mathcal{V}' = \{v'_1, \ldots, v'_m\} \subset E$ a forma φ será representada pela matriz $\mathbf{b}' = \mathbf{p}^T \mathbf{b} \mathbf{p}$, onde $p \in M(m \times m)$ é a matriz de passagem de $\mathcal{V}$ para $\mathcal{V}'$. Como $\mathbf{p}$ e $\mathbf{p}^T$ são invertíveis, é claro que $\mathbf{b}$ e $\mathbf{b}' = \mathbf{p}^T \mathbf{b} \mathbf{p}$ têm o mesmo posto.

Em particular, se existir uma base $\mathcal{V} = \{v_1, \ldots, v_m\} \subset E$ tal que

$$\varphi(v) = \lambda_1 x_1^2 + \cdots + \lambda_r x_r^2 \quad (\text{com } \lambda_1 \neq 0, \ldots, \lambda_r \neq 0)$$

para todo $v = \Sigma x_i v_i$, então a matriz de φ na base $\mathcal{V}$ tem $\lambda_1, \ldots, \lambda_r$ na diagonal e os demais elementos iguais a zero, logo o posto de φ, igual ao posto desta matriz, é r.

Se E tem um produto interno e $\varphi(v) = \langle v, Bv \rangle$ então o posto da forma quadrática φ é igual ao posto do operador auto-adjunto $B \colon E \to E$.

Teorema 18.4. (Lei da inércia, de Sylvester.) *Se existir uma base $\mathcal{V} = \{v_1, \ldots, v_m\} \subset E$ tal que, para todo $v = \Sigma x_j v_j \in E$ se tem*

$$\varphi(v) = -x_1^2 - \cdots - x_i^2 + x_{i+1}^2 + \cdots + x_r^2$$

então a forma quadrática φ tem índice i e posto r.

Demonstração: Já vimos acima que o posto de φ é r. Quanto ao índice, seja $F \subset E$ um subespaço vetorial, restrito ao qual φ é negativa. Mostremos que $\dim F \leq i$. Com efeito, se o vetor não-nulo

$$v = \sum_{j=1}^{m} x_j v_j$$

pertencer a F então

$$-x_1^2 - \cdots - x_i^2 + x_{i+1}^2 + \cdots + x_r^2 < 0,$$

logo

$$x_1^2 + \cdots + x_i^2 > 0$$

e daí $(x_1, \ldots, x_i) \neq 0$. Isto mostra que a transformação linear $A \colon F \to \mathbb{R}^i$, definida por

$$Av = A\left(\sum_{j=1}^{m} x_j v_j\right) = (x_1, \ldots, x_i),$$

é injetiva, portanto $\dim F \leq i$. Para finalizar, observamos que o subespaço gerado por $v_1, \ldots, v_i$ tem dimensão i e, restrita a ele,

$$\varphi(v) = -x_1^2 - \cdots - x_i^2$$

é negativa, Logo i é a maior dimensão de um subespaço de E, restrita ao qual φ é negativa. □

Usualmente, a Lei da Inércia de Sylvester é enunciada sob a seguinte forma equivalente: seja qual for a base $\mathcal{V} = \{v_1, \ldots, v_m\} \subset E$ tal que

$$\varphi(v) = -x_1^2 - \cdots - x_i^2 + x_{i+1}^2 + \cdots + x_r^2 \qquad (*)$$

para $v = \Sigma x_j v_j$, os números i e r são os mesmos.

Apresentaremos a seguir o *método de Lagrange* (de completar o quadrado) para obter uma base no espaço vetorial E, relativamente à qual uma dada forma quadrática $\varphi\colon E \to \mathbb{R}$ tenha uma matriz diagonal cujos elementos não-nulos são iguais a 1 ou -1. Noutras palavras, nessa base, φ tem uma expressão do tipo (*) acima.

Comecemos com uma base qualquer $\mathcal{U} = \{u_1, \ldots, u_m\} \subset E$, na qual se tem

$$\varphi(v) = \sum_{i,j} b_{ij} x_i x_j, \qquad (**)$$

para $v = x_1u_1 + \cdots + x_mu_m$. Se $\mathbf{a} = [a_{ij}] \in M(m \times m)$ é a matriz de passagem da base $\mathcal{U}$ para a base $\mathcal{V} = \{v_1, \ldots, v_m\}$, a nova expressão

$$\varphi(v) = \sum_{i,j} c_{ij} y_i y_j,$$

para $v = y_1v_1 + \cdots + y_mv_m$, se obtém simplesmente efetuando em (**) a substituição, ou mudança de variáveis,

$$x_i = \sum_k a_{ik} y_k, \quad x_j = \sum_k a_{jk} y_k.$$

De agora em diante, mencionaremos apenas cada mudança de variáveis, deixando implícito que ela corresponde à passagem de uma base para outra pois, em cada caso, a matriz $\mathbf{a} = [a_{ij}]$ [ou o que é o mesmo, o operador $(y_1, \ldots, y_m) \mapsto (x_1, \ldots, x_m)$] é invertível.

A partir da expressão

$$\varphi(v) = \sum_{i,j} b_{ij} x_i x_j\,,$$

faremos, se necessário, uma mudança de variáveis para assegurar que algum elemento b_{ii} da diagonal seja diferente de zero. (Estamos supondo que φ não seja identicamente nula, do contrário nada haveria a fazer.) Pois se todos os elementos da diagonal fossem iguais a zero, escolheríamos um $b_{rs} \neq 0$, com $r \neq s$ e faríamos a mudança de variável $x_r = y_r + y_s$, $x_s = y_r - y_s$, $x_i = y_i$ (se $i \notin \{r, s\}$). Então a parcela $2b_{rs}x_rx_s$ da soma $\varphi(v)$ se transformaria em $2b_{rs}y_r^2 - 2b_{rs}y_s^2$ e agora, nas novas variáveis y_i, a expressão de $\varphi(v)$ contém ao menos dois termos não nulos, $c_{rr}y_r^2 = 2b_{rs}y_r^2$ e $c_{ss}y_s^2 = -2b_{rs}y_s^2$, os quais não se cancelam com nenhuma das outras parcelas.

Voltemos à expressão $\varphi(v) = \Sigma b_{ij} x_i x_j$, onde podemos agora supor que algum $b_{ii} \neq 0$. Mudando, se necessário, a numeração das variáveis (o que equivale a mudar a ordem dos vetores da base), podemos admitir que $b_{11} \neq 0$. Então escrevemos

$$\varphi(v) = b_{11}\left(x_1^2 + 2x_1 \cdot \sum_{j\geq 2} c_j x_j\right) + \psi(v'),$$

onde $c_j = b_{1j}/b_{11}$ e $\psi(v')$ depende apenas de

$$v' = \sum_{j\geq 2} x_j u_j,$$

ou seja, ψ é uma forma quadrática definida no subespaço $F \subset E$, de dimensão $m-1$, gerado por $u_2, \ldots, u_m$. A expressão dentro dos parênteses acima é do tipo $a^2 + 2ab$, logo é igual a $(a+b)^2 - b^2$, onde

$$a = x_1 \quad \text{e} \quad b = \sum_{j\geq 2} c_j x_j.$$

Assim, a mudança de variáveis

$$z_1 = x_1 + \sum_{j\geq 2} c_j x_j, \quad z_2 = x_2, \ldots, z_m = x_m$$

ou, equivalentemente,

$$x_1 = z_1 - \sum_{j\geq 2} c_j z_j, \quad x_2 = z_2, \ldots, x_m = z_m\,,$$

mostra que a expressão dentro dos parênteses acima é igual a

$$z_1^2 - \left(\sum_{j\geq 2} c_j z_j\right)^2,$$

portanto

$$\varphi(v) = b_{11} z_1^2 - b_{11}\left(\sum_{j\geq 2} c_j z_j\right)^2 + \psi(v')$$

ou seja, $\varphi(v) = b_{11}z_1^2 + \xi(v')$ onde, outra vez, ξ é uma forma quadrática no subespaço F, de dimensão $m-1$. Efetuando a nova mudança de variáveis

$$z_1 = \frac{y_1}{\sqrt{|b_{11}|}}, \quad z_j = y_j \quad (j \geq 2),$$

obtemos $\varphi(v) = \pm y_1^2 + \xi(v')$. Aplicando o mesmo método a ξ e prosseguindo analogamente, chegaremos no final a

$$\varphi(v) = \sum \pm y_i^2$$

ou, mais explicitamente:

$$\varphi(v) = -y_1^2 - \cdots - y_i^2 + y_{i+1}^2 + \cdots + y_r^2,$$

onde i é o índice e r o posto da forma quadrática φ. Os números y_j são as coordenadas do vetor v na base de E obtida após a última mudança de coordenadas.

Exemplo 18.2. Seja $\varphi \colon \mathbb{R}^3 \to \mathbb{R}$ a forma quadrática dada por

$$\varphi(x,y,z) = 2x^2 + 2y^2 - 2z^2 + 4xy - 4xz + 8yz.$$

Completando o quadrado, vemos que a soma das parcelas que contêm o fator x se escreve como

$$\begin{aligned} 2x^2 + 4xy - 4xz &= 2[x^2 + 2x(y-z)] \\ &= 2[x^2 + 2x(y-z) + (y-z)^2] - 2(y-z)^2 \\ &= 2(x+y-z)^2 - 2(y-z)^2. \end{aligned}$$

A mudança de variáveis $s = x + y - z$ nos dá então

$$\begin{aligned} \varphi(x,y,z) &= 2s^2 - 2(y-z)^2 + 2y^2 - 2z^2 + 8yz \\ &= 2s^2 - 4z^2 + 12yz. \end{aligned}$$

Novamente completamos o quadrado, agora na soma das parcelas que contêm z, e obtemos

$$\begin{aligned} -4z^2 + 12yz &= -4\left(z^2 - 2z \cdot \frac{3}{2}y\right) \\ &= -4\left(z^2 - 2z \cdot \frac{3}{2}y + \frac{9}{4}y^2\right) + 9y^2 \\ &= -4\left(z - \frac{3}{2}y\right)^2 + 9y^2. \end{aligned}$$

A mudança de variáveis $t = z - \frac{3}{2}y$ nos dá portanto

$$\varphi(x, y, z) = 2s^2 - 4t^2 + 9y^2.$$

Isto já nos diz que a forma quadrática φ tem posto 3 e índice 1, logo é indefinida.

Se quisermos, podemos fazer a mudança de variáveis

$$s = \frac{y_1}{\sqrt{2}}, \quad t = \frac{y_2}{2}, \quad y = \frac{y_3}{3}$$

e teremos

$$\varphi(x, y, z) = y_1^2 - y_2^2 + y_3^2.$$

Evidentemente, não há dificuldade em obter a expressão das variáveis y_1, y_2 e y_3 em função de x, y, z e vice-versa. Mas, para efeito de conhecer o índice e o posto de φ, isto é desnecessário.

Quádricas

O estudo das formas quadráticas tem uma aplicação interessante à Geometria Analítica n-dimensional, que apresentaremos brevemente aqui.

Um subconjunto $\Sigma \subset \mathbb{R}^n$ chama-se uma *quádrica central* quando existe uma forma quadrática $\varphi \colon \mathbb{R}^n \to \mathbb{R}$ tal que Σ é definido pela equação $\varphi(v) = 1$.

Se

$$\varphi(v) = \sum_{i,j} a_{ij}\, x_i x_j$$

para $v = (x_1, \ldots, x_n)$, isto significa que Σ é o conjunto dos pontos $(x_1, \ldots, x_n) \in \mathbb{R}^n$ tais que

$$\sum_{i,j=1}^{n} a_{ij}\, x_i x_j = 1.$$

Segue-se imediatamente do Teorema 18.3 que, dada uma quádrica central $\Sigma \subset \mathbb{R}^n$, existe uma base ortonormal $\mathcal{U} = \{u_1, \ldots, u_n\}$ em $\mathbb{R}^n$, relativamente à qual a equação de Σ é

$$\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = 1,$$

onde $\lambda_1, \ldots, \lambda_n$ são os autovalores da matriz simétrica $[a_{ij}]$.

Noutras palavras, quando se exprime o vetor $v = y_1 u_1 + \cdots + y_n u_n$ como combinação linear dos elementos da base $\mathcal{U}$, tem-se $v \in \Sigma$ se, e somente se, suas coordenadas $y_1, \ldots, y_n$ satisfazem a equação

$$\sum \lambda_i y_i^2 = 1.$$

As retas que contêm os vetores u_i da base $\mathcal{U}$ chamam-se os *eixos principais* da quádrica central Σ.

Quando $n = 2$, Σ chama-se uma *cônica*. Portanto, a cônica definida em $\mathbb{R}^2$ pela equação

$$ax^2 + 2bxy + cy^2 = 1$$

pode, numa nova base ortonormal de $\mathbb{R}^2$, ser representada pela equação

$$\lambda s^2 + \mu t^2 = 1,$$

onde λ, μ são os autovalores da matriz $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$.

Segundo os sinais desses autovalores, as seguintes possibilidades podem ocorrer:

1º) Se $\lambda\mu < 0$ (isto é, $ac < b^2$) tem-se uma *hipérbole*.

2º) Se $\lambda > 0$ e $\mu > 0$ (isto é, $a > 0$ e $ac > b^2$) tem-se uma *elipse*.

3º) Se $\lambda < 0$ e $\mu < 0$ (isto é, $a < 0$ e $ac > b^2$) tem-se o conjunto vazio.

4º) Se $\lambda\mu = 0$ (isto é, $ac = b^2$) tem-se um par de retas paralelas, ou o conjunto vazio, conforme seja $a + c > 0$ ou $a + c \leq 0$.

Evidentemente, quando $\lambda = \mu > 0$ tem-se uma circunferência.

É claro que, em cada caso concreto, dada a equação $ax^2 + 2bxy + cy^2 = 1$, uma aplicação imediata do método de Lagrange (de completar o quadrado) permite escrever o primeiro membro sob a forma $b_1 y_1^2 + b_2 y_2^2$ e daí constatar, conforme os sinais de b_1 e b_2, se se trata de elipse, hipérbole ou um par de retas. Convém ter em conta, porém, que o método de Lagrange, embora eficiente (principalmente

em dimensões maiores do que 2, onde é muito difícil calcular os autovalores), não fornece bases ortonormais. Em particular, ele não permite distinguir uma elipse de uma circunferência.

Em dimensão 3, as superfícies quádricas centrais $\Sigma \subset \mathbb{R}^3$ têm, num sistema de coordenadas ortogonais conveniente, uma equação do tipo

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 = 1.$$

As possibilidades são as seguintes:

1º) Se $\lambda_1 > 0, \lambda_2 > 0$ e $\lambda_3 > 0$, Σ é um *elipsóide*.

2º) Se $\lambda_1 > 0, \lambda_2 > 0$ e $\lambda_3 < 0$, Σ é um *hiperbolóide de revolução*.

3º) Se $\lambda_1 > 0, \lambda_2 < 0$ e $\lambda_3 < 0$, Σ é um *hiperbolóide de duas folhas*.

4º) Se λ_1, λ_2 e λ_3 são ≤ 0, Σ é o conjunto vazio.

5º) Se $\lambda_3 = 0$, $\Sigma = C \times \mathbb{R} = \{(v, t); v \in C, t \in \mathbb{R}\}$, onde $C \subset \mathbb{R}^2$ é definido pela equação

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1.$$

Neste caso, Σ é um cilindro de base C.

Observe que, mudando a numeração das coordenadas se necessário, podemos sempre supor $\lambda_1 > 0$.

Há outro tipo mais geral de quádricas, além das centrais. São subconjuntos $\Sigma \subset \mathbb{R}^n$ definidos por uma equação do tipo

$$\sum_{i,j=1}^{n} a_{ij}\, x_i x_j + \sum_{i=1}^{n} b_i x_i = c.$$

A escolha de uma base ortonormal conveniente em $\mathbb{R}^n$ faz com que esta equação assuma a forma

$$\sum_{i=1}^{r} \lambda_i y_i^2 + \sum_{i=1}^{n} b_i' y_i = c,$$

onde r é o posto da matriz $[a_{ij}]$ e $\lambda_1 \neq 0, \ldots, \lambda_r \neq 0$.

Procurando eliminar o termo linear $\Sigma b_i' y_i$ mediante escolha de novas coordenadas, somos levados a efetuar uma translação. (Até

agora nossas mudanças de coordenadas eram feitas por meio de transformações lineares, logo preservavam a origem.) Introduzindo as novas coordenadas

$$\begin{aligned} z_i &= y_i + \frac{b'_i}{2\lambda_i} \quad (i = 1, \ldots, r), \\ z_{r+1} &= y_{r+1}, \\ &\vdots \\ z_n &= y_n, \end{aligned}$$

a equação acima se torna

$$\lambda_1 z_1^2 + \cdots + \lambda_r z_r^2 + b'_{r+1} z_{r+1} + \cdots + b'_n z_n = c',$$

na qual conseguimos eliminar os r primeiros termos lineares.

Pode ocorrer que os coeficientes b'_j sejam todos iguais a zero. Neste caso, a forma simplificada que buscamos para a equação da quádrica Σ é:

$$\lambda_1 z_1^2 + \cdots + \lambda_r z_r^2 = c'.$$

Se $r = n$, Σ é simplesmente a figura que resulta de uma quádrica central depois de uma translação. Se $r < n$ então as últimas $n - r$ coordenadas $z_{r+1}, \ldots, z_n$ podem ser tomadas arbitrariamente, enquanto as r primeiras definem uma quádrica em $\mathbb{R}^r$, logo Σ é um cilindro generalizado: produto cartesiano $\Sigma = \Sigma' \times \mathbb{R}^{n-r}$ onde Σ' é uma quádrica em $\mathbb{R}^r$.

Se algum dos números b'_j $(r + 1 \leq j \leq n)$ for $\neq 0$, introduziremos novas coordenadas $t_1, \ldots, t_n$ de tal modo que $t_1 = z_1, \ldots, t_r = z_r$ e

$$b'_{r+1} z_{r+1} + \cdots + b'_n z_n - c' = d t_{r+1}.$$

Primeiro fazemos uma translação de modo a eliminar c'. Para isso escolhemos um ponto $(z^o_{r+1}, \ldots, z^o_n)$ tal que

$$b'_{r+1} z^o_{r+1} + \cdots + b'_n z^o_n = c'$$

e escrevemos

$$s_1 = z_1, \ldots, s_r = z_r, \quad s_{r+1} = z_{r+1} - z^o_{r+1}, \ldots, s_n = z_n - z^o_n.$$

Então, nas coordenadas s_i, a equação da quádrica Σ se torna:

$$\lambda_1 s_1^2 + \cdots + \lambda_r s_r^2 + b'_{r+1} s_{r+1} + \cdots + b'_n s_n = 0.$$

No espaço $\mathbb{R}^{n-r}$, a expressão

$$b'_{r+1} s_{r+1} + \cdots + b'_n s_n$$

é o produto interno $\langle b, w \rangle$, onde

$$b = (b'_{r+1}, \ldots, b'_n) \quad \text{e} \quad w = (s_{r+1}, \ldots, s_n).$$

Escolhamos nesse espaço uma base ortonormal cujo primeiro elemento seja o vetor unitário $u = b/|b|$. Sejam $(t_{r+1}, \ldots, t_n)$ as coordenadas do vetor $w = (s_{r+1}, \ldots, s_n)$ nesta base. Então $t_{r+1} = \langle u, w \rangle$, logo $\langle b, w \rangle = |b| t_{r+1}$. Escrevendo $t_1 = s_1, \ldots, t_r = s_r$, vemos que, nas novas coordenadas t_i a equação da quádrica Σ assume a forma:

$$\lambda_1 t_1^2 + \cdots + \lambda_r t_r^2 + d t_{r+1} = 0,$$

onde $d = |b|$.

Observemos que todas as mudanças de coordenadas foram feitas mediante translações e transformações lineares ortogonais (ou, o que é o mesmo, escolhas de bases ortonormais). Podemos então concluir que, dada uma quádrica:

$$\Sigma : \sum_{i,j=1}^{n} a_{ij}\, x_i x_j + \sum_{i=1}^{n} b_i x_i = c \tag{*}$$

em $\mathbb{R}^n$, existe uma mudança de coordenadas

$$x_i = \sum_{j=1}^{n} m_{ij} t_j + k_i \quad (i = 1, \ldots, n)$$

tal que $\mathbf{m} = [m_{ij}]$ é uma matriz ortogonal e, efetuando essa transformação na equação (*) de Σ, ela se torna

$$\lambda_1 t_1^2 + \cdots + \lambda_r t_r^2 + d t_{r+1} = 0, \quad (d \neq 0)$$

ou

$$\lambda_1 t_1^2 + \cdots + \lambda_r t_r^2 = c',$$

onde r é o posto da matriz $[a_{ij}]$ e $\lambda_1, \ldots, \lambda_r$ são os seus autovalores não-nulos. No primeiro caso, supondo $r = n-1$, a quádrica Σ pode ser definida por

$$t_n = \alpha_1 t_1^2 + \cdots + \alpha_{n-1} t_{n-1}^2$$

(com $\alpha_i = -\lambda_i/d$) e chama-se um *parabolóide*.

Exercícios

18.1. Determine a matriz de cada uma das formas bilineares abaixo, relativamente à base especificada:

(a) $b\colon \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$, $b(u,v) = \langle u, v\rangle$, base $\{v_1, v_2, v_3, v_4\} \subset \mathbb{R}^4$, onde $v_1 = (-2, 0, 3, 1)$, $v_2 = (1, 2, 1, -1)$, $v_3 = (0, 1, 2, -1)$, $v_4 = (1, 2, 3, 1)$.

(b) $b\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $b(u,v) = \langle Au, v\rangle$, $A \in \mathcal{L}(\mathbb{R}^n)$, base canônica de $\mathbb{R}^n$.

(c) $b\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $b(u,v) = \langle Au, Bv\rangle$, $A, B \in \mathcal{L}(\mathbb{R}^n)$, base canônica de $\mathbb{R}^n$.

(d) $b\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $b(u,v) = \langle u, a\rangle \cdot \langle v, b\rangle$, $a, b \in \mathbb{R}^n$, base canônica de $\mathbb{R}^n$.

18.2. Seja $\varphi\colon \mathbb{R}^3 \to \mathbb{R}$ a forma quadrática dada por $\varphi(x,y,z) = x^2 + y^2 - z^2 + 2xy - 3xz + yz$. Qual é a matriz de φ na base $\{u, v, w\} \subset \mathbb{R}^3$, onde $u = (3, 0, 1)$, $v = (1, -1, 2)$, $w = (2, 1, 2)$?

18.3. Prove que o conjunto das formas quadráticas no espaço vetorial E é um subespaço vetorial $Q(E) \subset \mathcal{F}(E; \mathbb{R})$. Se $\dim E = n$, prove que $\dim Q(E) = n(n+1)/2$.

18.4. Assinale V(erdadeiro) ou F(also):

( ) O conjunto das formas quadráticas de índice i no espaço vetorial E é um cone convexo em $\mathcal{F}(E; \mathbb{R})$.

( ) A matriz do operador $B\colon E \to E$ na base $\mathcal{U}$ é igual à matriz da forma bilinear $b(u,v) = \langle u, Bv\rangle$ na mesma base.

( ) As formas bilineares b e b^*, que correspondem aos operadores B e B^* segundo o Teorema 18.2, definem a mesma forma quadrática $\varphi(v) = b(v, v) = b^*(v, v)$.

( ) O operador auto-adjunto $B \colon E \to E$ e a forma bilinear simétrica $b \colon E \times E \to \mathbb{R}$, que lhe corresponde conforme o Teorema 18.2, têm a mesma matriz relativamente a qualquer base de E.

18.5. A matriz de uma forma bilinear $b \colon E \times E$ na base $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ é dada por $b_{ij} = b(u_i, u_j)$. Supondo conhecida a forma quadrática $\varphi \colon E \to \mathbb{R}$, como se pode obter a matriz de φ na base $\mathcal{U}$ a partir dos seus valores $\varphi(v)$, $v \in E$?

18.6. Sejam $f, g \colon E \to \mathbb{R}$ funcionais lineares. Prove que a forma bilinear anti-simétrica $f \wedge g \colon E \to E$, definida por $(f \wedge g)(u, v) = f(u)g(v) - f(v)g(u)$, é identicamente nula se, e somente se, um dos funcionais dados é múltiplo do outro.

18.7. Seja $b \colon E \times E \to \mathbb{R}$ uma forma bilinear anti-simétrica no espaço vetorial E, de dimensão finita. Prove que as seguintes afirmações sobre b são equivalentes:

(1) Tem-se $b = f \wedge g$, onde $f, g \in E^*$.

(2) Existe uma base $\mathcal{V} = \{v_1, \ldots, v_n\} \subset E$ tal que $b(v_i, v_j) = 0$ se $\{i, j\} \neq \{1, 2\}$.

Conclua que toda forma bilinear em $\mathbb{R}^3$ é do tipo $b = f \wedge g$. E em $\mathbb{R}^4$?

18.8. Seja $b \colon \mathbb{R}^5 \times \mathbb{R}^5 \to \mathbb{R}$ uma forma bilinear anti-simétrica. Prove que existem números α, β e uma base $\{v_1, v_2, v_3, v_4, v_5\} \subset \mathbb{R}^5$ tais que $b(v_1, v_2) = -b(v_2, v_1) = \alpha$, $b(v_3, v_4) = -b(v_4, v_3) = \beta$ e $b(v_i, v_j) = 0$ nos demais casos. Conclua que existem funcionais lineares $f_1, f_2, f_3, f_4 \colon \mathbb{R}^5 \to \mathbb{R}$ tais que $b = f_1 \wedge f_2 + f_3 \wedge f_4$.

18.9. Sejam E, F espaços vetoriais de dimensão finita, com produto interno. Estabeleça um isomorfismo natural entre $\mathcal{B}(E \times F)$ e $\mathcal{L}(E; F)$. Determine a relação entre as matrizes de uma forma bilinear e da transformação linear que lhe corresponde.

18.10. Dados os funcionais lineares $f, g \colon E \to \mathbb{R}$, considere a forma bilinear $f \otimes g \colon E \times E \to \mathbb{R}$ (produto tensorial de f e g) definida por $(f \otimes g)(u, v) = f(u) \cdot g(v)$. Prove as seguintes afirmações:

(a) $f \otimes g = 0 \Rightarrow f = 0$ ou $g = 0$.

(b) $f \otimes g = g \otimes f \Leftrightarrow$ um dos funcionais f, g é múltiplo do outro.

(c) Se $\{f_1 \ldots, f_n\} \subset E^*$ é uma base então as formas bilineares $f_i \otimes f_j$ constituem uma base de $\mathcal{B}(E \times E)$.

(d) Nas condições do item (c), as formas $f_i \bullet f_j = f_i \otimes f_j + f_j \otimes f_i$, com $i \leq j$, constituem uma base para o espaço das formas bilineares simétricas.

(e) Com a notação de (c), as formas $f_i \wedge f_j = f_i \otimes f_j - f_j \otimes f_i$, com $i < j$, constituem uma base para o espaço das formas bilineares anti-simétricas.

18.11. Dada a forma quadrática $\varphi \colon \mathbb{R}^2 \to \mathbb{R}$, com $\varphi(x, y) = 2x^2 - y^2 + 3xy$, ache uma base ortonormal $\{u, v\} \subset \mathbb{R}^2$ e números λ, μ tais que, para todo vetor $w = su + tv \in \mathbb{R}^2$ se tenha $\varphi(w) = \lambda s^2 + \mu t^2$.

18.12. Dada a matriz simétrica

$$\mathbf{b} = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix},$$

ache uma matriz ortogonal $\mathbf{p}$ tal que $\mathbf{p}^T\mathbf{bp}$ seja uma matriz diagonal. A partir daí obtenha uma base ortonormal $\{u, v\} \subset \mathbb{R}^2$ tal que a forma quadrática $\varphi \colon \mathbb{R}^2 \to \mathbb{R}$, com $\varphi(x, y) = 4x^2 + 2xy + 2y^2$ se escreva como $\varphi(w) = \lambda s^2 + \mu t^2$ para todo vetor $w = su + tv$.

18.13. Dadas as formas quadráticas abaixo, use o método de Lagrange para exprimi-las como somas e diferenças de quadrados e, a partir daí, determine o índice e o posto de cada uma delas:

(a) $\varphi(x, y) = x^2 + 9y^2 + 6xy$

(b) $\psi(x, y) = x^2 + 8y^2 + 6xy$

(c) $\xi(x, y, z) = x^2 + 2xy + z^2 - 3xz + y^2 - 2yz$

(d) $\zeta(x, y, z) = 8x^2 + 36yz - 6y^2 - 27z^2 - 24xy$

18.14. Dada a forma quadrática $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$, com $\varphi(x,y) = ax^2 + bxy + cy^2$, escreva-a como

$$\begin{aligned}\varphi(x,y) &= y^2\left[a\left(\frac{x}{y}\right)^2 + b\left(\frac{x}{y}\right) + c\right] \\ &= y^2[at^2 + bt + c], \quad t = x/y,\end{aligned}$$

e use o trinômio $at^2 + bt + c$ para obter condições sobre a, b, c que caracterizam se φ é positiva, negativa, indefinida, não-negativa ou não-positiva.

18.15. Seja $A\colon E \to F$ uma transformação linear invertível. Para toda forma quadrática $\psi\colon F \to \mathbb{R}$, prove que $\varphi = \psi \circ A\colon E \to \mathbb{R}$ é uma forma quadrática em E, com o mesmo índice e o mesmo posto que ψ.

18.16. Prove que todo operador linear invertível $A\colon \mathbb{R}^n \to \mathbb{R}^n$ transforma uma quádrica $\Sigma \subset \mathbb{R}^n$ noutra quádrica de mesmo tipo.

19

Determinantes

O determinante surgiu inicialmente nas fórmulas que exprimem a solução de um sistema determinado de n *equações lineares a* n *incógnitas. Posteriormente, ele foi identificado como a área de um paralelogramo ou o volume de um paralelepípedo e depois, de forma definitiva, como a função multilinear alternada da qual todas as outras se deduzem.*

Tradicionalmente se tem o determinante de uma matriz quadrada, ou de uma lista de vetores, que são as colunas (ou linhas) dessa matriz. Em Álgebra Linear, é mais apropriado falar do determinante de um operador linear. Seu interesse aqui decorre principalmente do fato de que um operador é invertível se, e somente se, seu determinante é diferente de zero. A partir daí, o determinante fornece um polinômio de grau n *cujas raízes reais são os autovalores de um operador num espaço vetorial* n*-dimensional (polinômio característico), polinômio esse que estudaremos na seção seguinte.*

Faremos aqui uma breve apresentação do conceito de determinante e de suas principais propriedades. Nosso ponto de partida são as funções multilineares, mais especificamente as alternadas.

Em toda esta seção, E representa um espaço vetorial de dimensão n, no qual todas as bases que tomaremos serão ordenadas.

Uma *forma* r*-linear* no espaço vetorial E é uma função

$$f \colon E \times \cdots \times E \to \mathbb{R},$$

definida no produto cartesiano $E \times \cdots \times E = E^r$, que é linear em cada uma das suas variáveis, isto é:

$$f(v_1, \ldots, v_i + v_i', \ldots, v_r) = f(v_1, \ldots, v_i, \ldots, v_r) + f(v_1, \ldots, v_i', \ldots, v_r),$$

$$f(v_1, \ldots, \alpha v_i, \ldots, v_r) = \alpha f(v_1, \ldots, v_i, \ldots, v_r),$$

para $v_1, \ldots, v_i, v_i' \ldots, v_r \in E$ e $\alpha \in \mathbb{R}$ quaisquer.

Se f é r-linear e um dos vetores $v_1, \ldots, v_r$ é igual a zero então $f(v_1, \ldots, v_r) = 0$, pois

$$f(\ldots, 0, \ldots) = f(\ldots, 0.0, \ldots) = 0.f(\ldots, 0, \ldots) = 0.$$

Teorema 19.1. *Seja* $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ *uma base. Para cada uma das* n^r *listas ordenadas* $J = (j_1, \ldots, j_r)$ *de inteiros compreendidos entre* 1 *e* n*, fixemos um número real* $a_J = a_{j_1 j_2 \ldots j_r}$*. Existe uma, e somente uma, forma* r*-linear* $f \colon E \times \cdots \times E \to \mathbb{R}$ *tal que* $f(u_{j_1}, \ldots, u_{j_r}) = a_J$ *para todo* $J = (j_1, \ldots, j_r)$.

Noutras palavras, uma forma r-linear $f \colon E^r \to \mathbb{R}$ fica inteiramente determinada quando se conhecem seus n^r valores $f(u_{j_1}, \ldots, u_{j_r})$ nas listas de elementos básicos, e esses valores podem ser escolhidos arbitrariamente.

Demonstração: Para facilitar a escrita, tomemos $r = 3$. O caso geral se trata igualmente. Suponhamos dado, para cada lista (i, j, k) de números naturais compreendidos entre 1 e n, um número real a_{ijk}. Definamos $f \colon E \times E \times E \to \mathbb{R}$ pondo

$$f(u, v, w) = \sum_{i,j,k=1}^{n} a_{ijk} x_i y_j z_k,$$

se

$$u = \sum x_i u_i, \quad v = \sum y_j u_j \quad \text{e} \quad w = \sum z_k u_k.$$

Verifica-se facilmente que f, assim definida, é trilinear e que $f(u_i, u_j, u_k) = a_{ijk}$. Além disso, se $g \colon E \times E \times E \to \mathbb{R}$ é trilinear e $g(u_i, u_j, u_k) = a_{ijk}$ para $i, j, k = 1, \ldots, n$ quaisquer, então para

$$u = \sum x_i u_i, \quad v = \sum y_j u_j \quad \text{e} \quad w = \sum z_k u_k.$$

arbitrários, tem-se

$$\begin{aligned} g(u,v,w) &= g(\sum x_i u_i, \sum y_j u_j, \sum z_k u_k) \\ &= \sum x_i y_j z_k \, g(u_i,u_j,u_k) \\ &= \sum a_{ijk} x_i y_j z_k = f(u,v,w), \end{aligned}$$

logo $g = f$. □

Corolário. *O espaço vetorial* $\mathcal{L}_r(E;\mathbb{R})$ *das formas* r*-lineares* $f\colon E \times \cdots \times E \to \mathbb{R}$ *tem dimensão* n^r.

Com efeito, uma vez fixada a base $\mathcal{U} \subset E$, o Teorema 19.1 faz corresponder, biunivocamente, a cada $f \in \mathcal{L}_r(E;\mathbb{R})$ os n^r números a_J. Isto determina um isomorfismo entre $\mathcal{L}_r(E;\mathbb{R})$ e $\mathbb{R}^{n^r}$. □

Uma forma r-linear $f\colon E \times \cdots \times E \to \mathbb{R}$ chama-se *alternada* quando $f(v_1,\ldots,v_r) = 0$ sempre que a lista $(v_1,\ldots,v_r)$ tiver repetições, ou seja, quando

$$f(v_1,\ldots,v_{i-1},v,v_{i+1},\ldots,v_{j-1},v,v_{j+1},\ldots,v_r) = 0$$

para quaisquer $v, v_1,\ldots,v_r \in E$.

Teorema 19.2. *Uma forma* r*-linear* $f\colon E \times \cdots \times E \to \mathbb{R}$ *é alternada se, e somente se, é anti-simétrica, isto é,*

$$f(v_1,\ldots,v_i,\ldots,v_j,\ldots,v_r) = -f(v_1,\ldots,v_j,\ldots,v_i,\ldots,v_r)$$

para quaisquer $v_1,\ldots,v_r \in E$.

Demonstração: Por simplicidade, escrevamos

$$f(v_1,\ldots,u,\ldots,v,\ldots,v_r) = \varphi(u,v).$$

Então, f alternada implica $\varphi(u,u) = \varphi(v,v) = \varphi(u+v,u+v) = 0$, logo

$$\begin{aligned} 0 = \varphi(u+v,u+v) &= \varphi(u,u) + \varphi(u,v) + \varphi(v,u) + \varphi(v,v) \\ &= \varphi(u,v) + \varphi(v,u), \end{aligned}$$

portanto $\varphi(u,v) = -\varphi(v,u)$, de modo que f é anti-simétrica. Reciprocamente, se f é anti-simétrica então $\varphi(v,v) = -\varphi(v,v)$ logo $2\varphi(v,v) = 0$, $\varphi(v,v) = 0$ e f é alternada. □

Corolário. *Seja* $f\colon E\times\cdots\times E\to\mathbb{R}$ r*-linear alternada. Para toda permutação* σ *dos inteiros* $1,2,\ldots,r$*, e toda lista de vetores* $v_1,\ldots,v_r\in E$ *tem-se*

$$f(v_{\sigma(1)},\ldots,v_{\sigma(r)})=\varepsilon_\sigma\, f(v_1,\ldots,v_r),$$

isto é,

$$f(v_{\sigma(1)},\ldots,v_{\sigma(r)})=\pm f(v_1,\ldots,v_r),$$

onde o sinal é $+$ *se* σ *é uma permutação par e* $-$ *se* σ *é uma permutação ímpar.* (Vide Apêndice.)

Com efeito, passa-se da seqüência $(v_1,\ldots,v_r)$ para $(v_{\sigma(1)},\ldots,v_{\sigma(r)})$ mediante k transposições sucessivas, que correspondem a k mudanças de sinal no valor de f. Como $\varepsilon_\sigma=(-1)^k$, o corolário segue-se. $\square$

A notação $\mathcal{A}_r(E)$ indica o espaço vetorial das formas r-lineares alternadas $f\colon E\times\cdots\times E\to\mathbb{R}$.

Exemplo 19.1. Dados os funcionais lineares $f_1,\ldots,f_r\colon E\to\mathbb{R}$, a função $f\colon E\times\cdots\times E\to\mathbb{R}$, definida por

$$f(v_1,\ldots,v_r)=f_1(v_1)\cdot f_2(v_2)\cdots f_r(v_r)$$

é uma forma r-linear, chamada o *produto tensorial* dos funcionais lineares $f_1,\ldots,f_r$.

Exemplo 19.2. Todo funcional linear $f\colon E\to\mathbb{R}$ é uma forma 1-linear alternada, já que não é possível violar a condição de anti-simetria. Portanto $\mathcal{A}_1(E)=E^*$.

Exemplo 19.3. Qualquer aplicação r-linear $f\colon\mathbb{R}\times\cdots\times\mathbb{R}\to\mathbb{R}$ é do tipo $f(t_1,\ldots,t_r)=a\cdot t_1\cdot t_2\cdots t_r$, onde $a=f(1,\ldots,1)$. (Vide Teorema 19.1.) Quando $r>1$, f só pode ser alternada quando $a=0$. Logo $\mathcal{A}_r(\mathbb{R})=\{0\}$ se $r>1$.

Exemplo 19.4. A forma $f\colon\mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}$, definida por $f(u,v)=x_1y_2-x_2y_1$ quando $u=(x_1,x_2)$ e $v=(y_1,y_2)$, é bilinear alternada. Se $g\colon\mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}$ é qualquer outra forma bilinear alternada em $\mathbb{R}^2$ então, pondo $c=g(e_1,e_2)$, vem:

$$\begin{aligned}
g(u,v)&=g(x_1e_1+x_2e_2,y_1e_1+y_2e_2)\\
&=x_1y_1g(e_1,e_1)+x_1y_2g(e_1,e_2)\\
&\qquad+x_2y_1g(e_2,e_1)+x_2y_2g(e_2,e_2)\\
&=(x_1y_2-x_2y_1)g(e_1,e_2)\\
&=c\cdot(x_1y_2-x_2y_1)=c\cdot f(u,v)
\end{aligned}$$

pois $g(e_1, e_1) = g(e_2, e_2) = 0$ e $g(e_2, e_1) = -g(e_1, e_2)$. Isto mostra que toda forma bilinear alternada g é um múltiplo de f, logo dim $\mathcal{A}_2(\mathbb{R}^2)=1$.

Teorema 19.3. *Seja* $f\colon E \times \cdots \times E \to \mathbb{R}$ r*-linear alternada. Se os vetores* $v_1, \ldots, v_r$ *forem* L.D. *então* $f(v_1, \ldots, v_r) = 0$.

Demonstração: Isto é claro se $v_1 = 0$. Caso contrário, um dos vetores, digamos v_i, é combinação linear dos anteriores:

$$v_i = \sum_{j<i} \alpha_j v_j.$$

Então

$$\begin{aligned} f(v_1, \ldots, v_r) &= f(v_1, \ldots, \sum_{j<i} \alpha_j v_j, \ldots, v_r) \\ &= \sum_{j<i} \alpha_j \, f(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_r) = 0 \end{aligned}$$

pois f é alternada. □

Corolário 1. *Se existe uma forma* r*-linear alternada* $f\colon E \times \cdots \times E \to \mathbb{R}$ *tal que* $f(v_1, \ldots, v_r) \neq 0$ *então os vetores* $v_1, \ldots, v_r \in E$ *são linearmente independentes.*

Corolário 2. *Se* $r >$ dim E *então* $\mathcal{A}_r(E) = \{0\}$.

Com efeito, sendo $r >$ dim E, quaisquer r vetores $v_1, \ldots, v_r$ são linearmente dependentes, logo $f(v_1, \ldots, v_r) = 0$ para todo $f \in \mathcal{A}_r$ e todos $v_1, \ldots, v_r \in E$. □

A partir de agora, concentraremos nosso interesse nas formas n-lineares alternadas, $f \in \mathcal{A}_n(E)$, onde $n =$ dim E. O resultado fundamental é o teorema seguinte e, principalmente, seu Corolário 1.

Teorema 19.4. *Seja* $\mathcal{U} = \{u_1, \ldots, u_n\}$ *uma base de* E. *Dado arbitrariamente um número real* a, *existe uma única forma* n*-linear alternada* $f\colon E \times \cdots \times E \to \mathbb{R}$ *tal que* $f(u_1, \ldots, u_n) = a$.

Demonstração: Suponhamos inicialmente que exista $f \in \mathcal{A}_n(E)$ tal que $f(u_1, \ldots, u_n) = a$. Se $g \in \mathcal{A}_n(E)$ for tal que $g(u_1, \ldots, u_n)$ também é igual a a, então para toda lista de vetores $u_{i_1}, \ldots, u_{i_n} \in \mathcal{U}$ tem-se:

$$g(u_{i_1}, \ldots, u_{i_n}) = 0 = f(u_{i_1}, \ldots, u_{i_n}),$$

se houve repetições na lista;

$$
\begin{aligned}
g(u_{i_1},\ldots,u_{i_n}) &= g(u_{\sigma(1)},\ldots,u_{\sigma(n)})\\
&= \varepsilon_\sigma g(u_1,\ldots,u_n)\\
&= \varepsilon_\sigma a\\
&= \varepsilon_\sigma\, f(u_1,\ldots,u_n)\\
&= f(u_{i_1},\ldots,u_{i_n})
\end{aligned}
$$

se $(i_1,\ldots,i_n) = (\sigma(1),\ldots,\sigma(n))$ for uma permutação dos inteiros $(1,\ldots,n)$, isto é, se não houver repetições na lista. Segue-se então do Teorema 19.1 que $f = g$. Portanto a forma n-linear alternada f fica determinada por seu valor $f(u_1,\ldots,u_n)$.

Isto nos indica como obter uma forma n-linear alternada $f\colon E\times\cdots\times E \to \mathbb{R}$ tal que $f(u_1,\ldots,u_n) = a$. Para toda lista ordenada $(i_1,\ldots,i_n)$ de inteiros compreendidos entre 1 e n, poremos $f(u_{i_1},\ldots,u_{i_n}) = 0$ se houver repetições na lista, $f(u_{i_1},\ldots,u_{i_n}) = a$ se $(i_1,\ldots,i_n)$ for uma permutação par dos números $1,2,\ldots,n$ e $f(u_{i_1},\ldots,u_{i_n}) = -a$ se a lista $i_1,\ldots,i_n$ for uma permutação ímpar dos números $1,2,\ldots,n$. Pelo Teorema 19.1, há uma única forma n-linear $f\colon E\times\cdots\times E \to \mathbb{R}$ com estas propriedades. Resta provar que f é alternada. Com esse objetivo, tomamos um vetor $v = \Sigma x_i u_i$ qualquer. Então:

$$
\begin{aligned}
f(*v*v*) &= \sum_{i,j=1}^{n} x_i x_j\, f(*u_i*u_j*)\\
&= \sum_i x_i x_i f(*u_i*u_i*) + \sum_{i<j} x_i x_j f(*u_i*u_j*)\\
&\quad + \sum_{i>j} x_i x_j f(*u_i*u_j*)\\
&\overset{(1)}{=} 0 + \sum_{i<j} x_i x_j f(*u_i*u_j*) - \sum_{i>j} x_i x_j\, f(*u_j*u_i*)\\
&\overset{(2)}{=} \sum_{i<j} x_i x_j f(*u_i*u_j*) - \sum_{i<j} x_j x_i f(*u_i*u_j*)\\
&\overset{(3)}{=} \sum_{i<j} x_i x_j f(*u_i*u_j*) - \sum_{i<j} x_i x_j f(*u_i*u_j*) = 0.
\end{aligned}
$$

Acima, os asteriscos estão em lugar de $n-2$ vetores que permanecem fixos durante a argumentação. A igualdade (1) usa o fato de que $f(*u_i * u_j *)$ é uma função $(n-2)$-linear desses vetores, que satisfaz a condição $f(*u_i * u_j *) = -f(*u_j * u_i *)$ quando esses $n-2$ vetores pertencem à base $\mathcal{U}$ logo, pelo Teorema 19.1, esta igualdade vale em geral. Segue-se que $f(*u_i * u_i *) = 0$. Na igualdade (2) foram trocados os nomes dos índices mudos i, j. Finalmente, a igualdade (3) usa apenas que $x_i x_j = x_j x_i$. □

Corolário 1. *Se* dim $E = n$ *então* dim $\mathcal{A}_n(E) = 1$.

Com efeito, fixada a base $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$, existe $f \in \mathcal{A}_n(E)$ tal que $f(u_1, ..., u_n)=1$. Então, para toda $g\in\mathcal{A}_n(E)$, se $g(u_1, ..., u_n) = a$ tem-se também $(af)(u_1, \ldots, u_n) = a$, logo $g = af$. Portanto, $\{f\}$ é uma base de $\mathcal{A}_n(E)$. □

Corolário 2. *Se* $\{u_1, \ldots, u_n\} \subset E$ *é uma base e* $0 \neq f \in \mathcal{A}_n(E)$ *então* $f(u_1, \ldots, u_n) \neq 0$.

Com efeito, pelo Teorema 19.4 existe $f_o \in \mathcal{A}_n(E)$ tal que $f_o(u_1, \ldots, u_n) = 1$. Pelo Corolário 1, $f = af_o$, com $a \neq 0$. Logo $f(u_1, \ldots, u_n) = af_o(u_1, \ldots, u_n) = a \neq 0$. □

Toda transformação linear $A\colon E \to F$ induz uma transformação linear $A^\#\colon \mathcal{A}_n(F) \to \mathcal{A}_n(E)$, a qual faz corresponder a cada forma n-linear alternada $f\colon F \times \cdots \times F \to \mathbb{R}$ a nova forma $A^\# f\colon E \times \cdots \times E \to \mathbb{R}$, definida por

$$(A^\# f)(v_1, \ldots, v_n) = f(Av_1, \ldots, Av_n),$$

onde $v_1, \ldots, v_n \in E$. É fácil verificar que realmente $A^\# f \in \mathcal{A}_n(E)$, que $(BA)^\# = A^\# B^\#$ e $I^\# = I$. Se A for um isomorfismo, $A^\#$ também será, com $(A^\#)^{-1} = (A^{-1})^\#$.

Seja agora $A\colon E \to E$ um operador linear.

Como dim $\mathcal{A}_n(E) = 1$, o operador linear $A^\#\colon \mathcal{A}_n(E) \to \mathcal{A}_n(E)$ consiste na multiplicação por um número real, que chamaremos o *determinante* do operador $A\colon E \to E$ e indicaremos com a notação det A. Assim, por definição, $A^\# f = \det A \cdot f$, isto é,

$$f(Av_1, \ldots, Av_n) = \det A \cdot f(v_1, \ldots, v_n)$$

para toda $f\colon E \times \cdots \times E \to \mathbb{R}$ n-linear alternada e quaisquer $v_1, \ldots, v_n \in E$.

Isto define o determinante de um operador de modo intrínseco, sem recurso a bases ou matrizes. A seguir mostraremos como obter, a partir desta definição, as formas clássicas de apresentar o determinante. De imediato, veremos como é fácil provar duas propriedades cruciais do determinante.

Teorema 19.5. *Se* $A, B: E \to E$ *são operadores lineares então* $\det(BA) = \det B \cdot \det A$.

Demonstração: Sejam $\{v_1, \ldots, v_n\} \subset E$ uma base e $0 \neq f \in \mathcal{A}_n(E)$, logo $f(v_1, \ldots, v_n) \neq 0$. Então

$$\begin{aligned}\det(BA).f(v_1, \ldots, v_n) &= f(BAv_1, \ldots, BAv_n) \\ &= \det B.f(Av_1, \ldots, Av_n) \\ &= \det B.\det A.f(v_1, \ldots, v_n),\end{aligned}$$

portanto $\det(BA) = \det B \cdot \det A$. □

Corolário. *Se* $A: E \to E$ *é um operador não-negativo então* $\det A \geq 0$.

Com efeito, existe $B: E \to E$ tal que $A = B^2$, logo $\det A = (\det B)^2 \geq 0$. □

O teorema abaixo mostra que $\det A > 0$ quando A é positivo.

Teorema 19.6. *O operador linear* $A: E \to E$ *é invertível se, e somente se,* $\det A \neq 0$. *No caso afirmativo, tem-se* $\det(A^{-1}) = (\det A)^{-1}$.

Demonstração: Se existe A^{-1}, então de $AA^{-1} = I$ resulta

$$\det A.\det(A^{-1}) = \det(AA^{-1}) = \det I = 1,$$

logo $\det A \neq 0$ e $\det(A^{-1}) = (\det A)^{-1}$. Reciprocamente, se $\det A \neq 0$ então, tomando uma base $\{v_1, \ldots, v_n\} \subset E$ e uma forma n-linear alternada não-nula $f: E \times \cdots \times E \to \mathbb{R}$ temos

$$f(Av_1, \ldots, Av_n) = \det A.f(v_1, \ldots, v_n) \neq 0,$$

logo, pelo Corolário 1 do Teorema 19.3, os vetores $Av_1, \ldots, Av_n$ constituem uma base de E. Assim, A é invertível. □

Corolário. *O sistema de equações lineares* $Ax = b$, *com* $A \in \mathcal{L}(\mathbb{R}^n)$, *possui uma única solução se, e somente se,* $\det A \neq 0$.

Definamos, a seguir, o determinante de uma matriz quadrada. Dada $\mathbf{a} \in M(n \times n)$, escreveremos, de agora em diante, $\mathbf{a} = [v_1, \ldots, v_n]$ para significar que $v_1, \ldots, v_n$ são os vetores-coluna da matriz $\mathbf{a}$. Seja $A\colon \mathbb{R}^n \to \mathbb{R}^n$ o operador linear cuja matriz na base canônica de $\mathbb{R}^n$ é $\mathbf{a}$, ou seja, $Ae_1 = v_1, \ldots, Ae_n = v_n$.

Por definição, o *determinante da matriz* $\mathbf{a}$ é igual a det A.

Assim, se $f_o \in \mathcal{A}_n(\mathbb{R}^n)$ é a forma n-linear alternada tal que $f_o(e_1, \ldots, e_n) = 1$ (vide Teorema 19.4) então

$$\det \mathbf{a} = \det A = \det A . f_o(e_1, \ldots, e_n) = f_o(Ae_1, \ldots, Ae_n),$$

ou seja:

$$\det \mathbf{a} = f_o(v_1, \ldots, v_n).$$

Portanto $\det\colon M(n \times n) \to \mathbb{R}$ *é a única função* n*-linear alternada das colunas da matriz* $\mathbf{a} = [v_1, \ldots, v_n]$ *que assume o valor* 1 *na matriz identidade* $\mathbf{I}_n$.

Para uma qualquer função n-linear alternada $f(\mathbf{a}) = f(v_1, \ldots, v_n)$ das colunas da matriz $\mathbf{a} = [v_1, \ldots, v_n]$, tem-se evidentemente $f(\mathbf{a}) = f(v_1, \ldots, v_n) = c \cdot \det \mathbf{a}$, onde $c = f(\mathbf{I}_n) = f(e_1, \ldots, e_n)$.

Escrevemos $\det[v_1, \ldots, v_n]$ para indicar o determinante da matriz cujas colunas são os vetores $v_1, \ldots, v_n \in \mathbb{R}^n$. A afirmação acima destacada é a caracterização axiomática dos determinantes feita por Weierstrass. Segundo ela, todas as propriedades dos determinantes podem, em princípio, ser deduzidas das seguintes:

1) $\det[\ldots, v_i + w_i, \ldots] = \det[\ldots, v_i, \ldots] + \det[\ldots, w_i, \ldots]$;

2) $\det[\ldots, \alpha v_i, \ldots] = \alpha \cdot \det[\ldots, v_i, \ldots]$;

3) $\det[\ldots, v_i, \ldots, v_j, \ldots] = -\det[\ldots, v_j, \ldots, v_i, \ldots]$;

4) $\det[e_1, \ldots, e_n] = 1$.

   Por exemplo, vale:

5) *O determinante de uma matriz não se altera quando se soma a uma de suas colunas uma combinação linear das demais.*

Seja

$$w = \sum_{j \neq i} \alpha_j v_j.$$

A afirmação acima significa que

$$\det[v_1,\ldots,v_i+w,\ldots,v_n]=\det[v_1,\ldots,v_i,\ldots,v_n],$$

o que é claro pois a segunda parcela do segundo membro abaixo é zero, pelo Teorema 19.3:

$$\det[\ldots,v_i+w,\ldots,]=\det[\ldots,v_i,\ldots]+\det[\ldots,w,\ldots].$$

Como aplicação de 5) temos o

Exemplo 19.5. O determinante de uma matriz triangular é igual ao produto dos elementos de sua diagonal. Com efeito, seja $\mathbf{t}=[t_{ij}]\in M(n\times n)$ triangular superior. Escrevendo $\mathbf{t}=[v_1,\ldots,v_n]$, temos

$$\begin{aligned}
v_1 &= t_{11}e_1,\\
v_2 &= t_{12}e_1+t_{22}e_2,\\
v_3 &= t_{13}e_1+t_{23}e_2+t_{33}e_3,\ \text{etc.}
\end{aligned}$$

Portanto

$$\begin{aligned}
\det\mathbf{t} &= \det[t_{11}e_1,v_2,\ldots,v_n]\\
&= t_{11}\det[e_1,t_{12}e_1+t_{22}e_2,v_3,\ldots]\\
&= t_{11}t_{22}\det[e_1,e_2,t_{13}e_1+t_{23}e_2+t_{33}e_3,v_4,\ldots]\\
&= t_{11}t_{22}t_{33}\det[e_1,e_2,e_3,v_4,\ldots].
\end{aligned}$$

Prosseguindo analogamente, chegamos a

$$\det\mathbf{t}=t_{11}t_{22}\ldots t_{nn}.$$

Dos Teoremas 19.5 e 19.6 resulta que $\det(\mathbf{ba})=\det\mathbf{b}\cdot\det\mathbf{a}$ e que $\det\mathbf{a}\neq 0$ se, e somente se, existe $\mathbf{a}^{-1}$. No caso afirmativo, $\det(\mathbf{a}^{-1})=(\det\mathbf{a})^{-1}$.

Exemplo 19.6. (Regra de Cramer.) Seja $\mathbf{a}\in M(n\times n)$ uma matriz invertível. Dado $b\in\mathbb{R}^n$, indiquemos com o símbolo $\mathbf{a}[i;b]$ a matriz obtida de $\mathbf{a}$ quando se substitui sua i-ésima coluna por b. A solução do sistema linear $\mathbf{ax}=\mathbf{b}$, de n equações a n incógnitas é o vetor $x=(x_1,\ldots,x_n)$ cujas coordenadas são

$$x_i=\frac{\det\mathbf{a}[i;b]}{\det\mathbf{a}},\quad i=1,\ldots,n.$$

Com efeito, se $\mathbf{a} = [v_1, \ldots, v_n]$, a igualdade $\mathbf{ax} = \mathbf{b}$ significa que $b = x_1v_1 + \cdots + x_nv_n$. Assim, para cada i de 1 até n, tem-se

$$\begin{aligned}\det \mathbf{a}[i; b] &= \det[v_1, \ldots, b, \ldots, v_n] \\ &= \sum_{k=1}^{n} x_k \det[v_1, \ldots, v_k, \ldots, v_n] \\ &= x_i \cdot \det[v_1, \ldots, v_n] \\ &= x_i \cdot \det \mathbf{a}.\end{aligned}$$

pois, quando $k \neq i$, a matriz $[v_1, \ldots, v_k, \ldots, v_n]$, com v_k na i-ésima coluna, tem duas colunas iguais a v_k, logo seu determinante é zero. Segue-se então que $x_i = \det \mathbf{a}[i; b] / \det \mathbf{a}$.

Se $\mathbf{a}$ e $\mathbf{a}'$ são matrizes do mesmo operador A em relação a bases diferentes então $\mathbf{a}' = \mathbf{p}^{-1}\mathbf{ap}$, onde $\mathbf{p}$ é a matriz de passagem de uma base para a outra. Portanto

$$\det \mathbf{a}' = (\det \mathbf{p})^{-1} \cdot \det \mathbf{a} \cdot \det \mathbf{p} = \det \mathbf{a}.$$

O teorema seguinte completa esta observação, mostrando que o determinante de todas essas matrizes é igual a $\det A$.

Teorema 19.7. *O determinante do operador linear* $A\colon E \to E$ *é igual ao determinante de uma matriz de* A *numa base qualquer de* E.

Demonstração: Seja $\mathbf{a} = [a_{ij}] \in M(n \times n)$ a matriz de A numa base $\mathcal{U} \subset E$. Por definição, $\det \mathbf{a} = \det A_o$, onde $A_o\colon \mathbb{R}^n \to \mathbb{R}^n$ é o operador cuja matriz na base canônica de $\mathbb{R}^n$ é $\mathbf{a}$. Seja $\varphi\colon E \to \mathbb{R}^n$ o isomorfismo que transforma $\mathcal{U}$ na base canônica de $\mathbb{R}^n$. Para cada $u_j \in \mathcal{U}$, tem-se $Au_j = \sum_i a_{ij}u_i$, logo

$$\varphi(Au_j) = \sum_i a_{ij}\varphi(u_i) = \sum_i a_{ij}e_i = A_oe_j = A_o\varphi(u_j).$$

Segue-se que $\varphi \cdot A = A_o \cdot \varphi$, ou seja, $A_o = \varphi \cdot A \cdot \varphi^{-1}$. Portanto, para cada $f \in \mathcal{A}_n(\mathbb{R}^n)$ tem-se

$$\begin{aligned}\det \mathbf{a} \cdot f = A_o^{\#}f = (\varphi^{\#})^{-1}A^{\#}\varphi^{\#}f = (\varphi^{\#})^{-1} \cdot \det A \cdot \varphi^{\#}f = \\ = \det A \cdot (\varphi^{\#})^{-1}\, \varphi^{\#}f = \det A \cdot f.\end{aligned}$$

Assim, $\det \mathbf{a} = \det A$. □

Tradicionalmente, o determinante de uma matriz $\mathbf{a} = [a_{ij}] \in M(n \times n)$ é definido como a soma de $n!$ parcelas do tipo

$$\pm a_{1j_1} a_{2j_2} \cdots a_{nj_n}.$$

Essas parcelas são produtos de n fatores que pertencem a linhas e colunas diferentes de $\mathbf{a}$. A ordem em que os fatores são dispostos é a ordem crescente dos índices $1, 2, \ldots, n$ das linhas. Os segundos índices (das colunas) exibem uma permutação $\sigma = (j_1, \ldots, j_n)$ dos inteiros $1, 2, \ldots, n$. O sinal que precede cada parcela é $+$ ou $-$, conforme a permutação σ seja par ou ímpar respectivamente. Noutras palavras, a definição clássica do determinante é

$$\det \mathbf{a} = \sum_\sigma \varepsilon_\sigma \cdot a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{n\sigma(n)},$$

o somatório sendo estendido a todas as permutações σ dos inteiros $1, 2, \ldots, n$, com $\varepsilon_\sigma = 1$ se σ é par e $\varepsilon_\sigma = -1$ se σ é ímpar.

Para obter esta expressão, consideremos mais uma vez a forma $f_o \in \mathcal{A}_n(\mathbb{R}^n)$ tal que $f_o(e_1, \ldots, e_n) = 1$, logo $f_o(e_{\sigma(1)}, \ldots, e_{\sigma(n)}) = \varepsilon_\sigma$ para toda permutação σ dos inteiros $1, 2, \ldots, n$.

Dada a matriz $\mathbf{a} = [a_{ij}] = [v_1, \ldots, v_n] \in M(n \times n)$, temos

$$\begin{aligned} v_1 &= \sum_{i_1=1}^{n} a_{i_1 1} e_{i_1}, \\ v_2 &= \sum_{i_2=1}^{n} a_{i_2 2} e_{i_2}, \\ &\vdots \\ v_n &= \sum_{i_n=1}^{n} a_{i_n n} e_{i_n}, \end{aligned}$$

logo

$$\begin{aligned} \det \mathbf{a} &= f_o(v_1, \ldots, v_n) \\ &= f_o\left(\sum_{i_1=1}^{n} a_{i_1 1} e_{i_1}, \sum_{i_2=1}^{n} a_{i_2 2} e_{i_2}, \ldots, \sum_{i_n=1}^{n} a_{i_n n} e_{i_n}\right) \\ &= \sum_{i_1, \ldots, i_n=1}^{n} a_{i_1 1} a_{i_2 2} \cdot \ldots \cdot a_{i_n n} \cdot f_o(e_{i_1}, e_{i_2}, \ldots, e_{i_n}). \end{aligned}$$

Neste último somatório, são nulas as parcelas em que há repetições entre os índices $i_1, i_2, \ldots, i_n$, restando apenas aquelas em que

$$(i_1, i_2, \ldots, i_n) = (\rho(1), \rho(2), \ldots, \rho(n))$$

é uma permutação dos inteiros $1, 2, \ldots, n$. Neste caso,

$$a_{i_1 1} \cdot a_{i_2 2} \cdot \ldots \cdot a_{i_n n} = a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{n\sigma(n)},$$

onde $\sigma = \rho^{-1}$ e a parcela correspondente é igual a

$$a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{n\sigma(n)} f_0(e_{\rho(1)}, \ldots, e_{\rho(n)}).$$

Por conseguinte, levando em conta que $\varepsilon_\sigma = \varepsilon_\rho$, tem-se:

$$\det \mathbf{a} = \sum_\sigma \varepsilon_\sigma a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{n\sigma(n)}\,, \qquad (*)$$

o somatório sendo estendido a todas as permutações σ dos inteiros $1, 2, \ldots, n$.

Voltando à expressão anterior (*), podemos escrever

$$\det \mathbf{a} = \sum_\rho \varepsilon_\rho a_{\rho(1)1} \cdot a_{\rho(2)2} \ldots a_{\rho(n)n}\,.$$

Esta igualdade mostra que o determinante da matriz $\mathbf{a}$ é igual ao determinante da sua transposta $\mathbf{a}^T$.

Se $A \colon E \to E$ é um operador num espaço vetorial munido de produto interno e $\mathbf{a}$ é sua matriz numa base ortonormal então

$$\det A = \det \mathbf{a} = \det \mathbf{a}^T = \det A^*.$$

Resulta da igualdade $\det \mathbf{a} = \det \mathbf{a}^T$ que o determinante é uma função n-linear alternada das *linhas* da matriz $\mathbf{a} \in M(n \times n)$. (A única tal que $\det \mathbf{I}_n = 1$.) Como no caso de colunas, uma qualquer função n-linear alternada f das linhas da matriz $\mathbf{a}$ é da forma $f(\mathbf{a}) = c \cdot \det \mathbf{a}$, onde $c = f(\mathbf{I}_n) = f(e_1, \ldots, e_n)$.

Outra conseqüência da igualdade $\det A = \det A^*$ é que o determinante de um operador ortogonal A é igual a ± 1, pois

$$(\det A)^2 = \det A^* \cdot \det A = \det(A^*A) = \det I = 1.$$

O teorema seguinte utiliza, em sua demonstração, a caracterização do determinante de uma matriz como função multilinear alternada das linhas ou das colunas dessa matriz.

Teorema 19.8. *Se* $\mathbf{b} \in M(r \times r)$, $\mathbf{c} \in M(r \times (n-r))$, $\mathbf{0} \in M((n-r) \times r)$ *e* $\mathbf{d} \in M((n-r) \times (n-r))$ *então o determinante da matriz*

$$\mathbf{a} = \begin{bmatrix} \mathbf{b} & \mathbf{c} \\ \mathbf{0} & \mathbf{d} \end{bmatrix} \quad \in \quad M(n \times n)$$

é igual a $\det \mathbf{b} \cdot \det \mathbf{d}$.

Demonstração: Para cada $\mathbf{c}$ fixa,

$$f(\mathbf{b}, \mathbf{d}) = \det \begin{bmatrix} \mathbf{b} & \mathbf{c} \\ \mathbf{0} & \mathbf{d} \end{bmatrix}$$

é uma função r-linear alternada das colunas de $\mathbf{b}$ e $(n-r)$-linear alternada das linhas de $\mathbf{d}$, logo

$$\begin{aligned} \det \begin{bmatrix} \mathbf{b} & \mathbf{c} \\ \mathbf{0} & \mathbf{d} \end{bmatrix} &= \det \mathbf{b} \cdot f(\mathbf{I}_r, \mathbf{d}) \\ &= \det \mathbf{b} \cdot \det \mathbf{d} \cdot f(\mathbf{I}_r, \mathbf{I}_{n-r}) \\ &= \det \mathbf{b} \cdot \det \mathbf{d}, \end{aligned}$$

pois

$$f(\mathbf{I}_r, \mathbf{I}_{n-r}) = \det \begin{bmatrix} \mathbf{I}_r & \mathbf{c} \\ \mathbf{0} & \mathbf{I}_{n-r} \end{bmatrix} = 1,$$

já que o determinante de uma matriz triangular é o produto dos elementos da sua diagonal. □

O Teorema 19.8 implica imediatamente uma sua versão mais geral, onde se tem uma matriz triangular por blocos, por exemplo, do tipo

$$\mathbf{a} = \begin{bmatrix} \mathbf{b} & \mathbf{c} & \mathbf{d} \\ & \mathbf{e} & \mathbf{f} \\ & & \mathbf{g} \end{bmatrix}.$$

Aí, $\mathbf{b}$, $\mathbf{e}$, $\mathbf{g}$ são matrizes quadradas (de ordens possivelmente diferentes) e $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ têm o mesmo número de linhas, assim como $\mathbf{e}$, $\mathbf{f}$. Além disso, $\mathbf{d}$, $\mathbf{f}$, $\mathbf{g}$ têm o mesmo número de colunas, assim como $\mathbf{c}$, $\mathbf{e}$. Os lugares em branco são ocupados por zeros.

Nestas condições, $\det \mathbf{a} = \det \mathbf{b} \cdot \det \mathbf{e} \cdot \det \mathbf{g}$.

Dada $\mathbf{a} = [a_{ij}] \in M(n \times n)$, indicaremos com a notação $\mathbf{M}_{ij}$ a matriz de ordem $(n-1) \times (n-1)$ obtida mediante a omissão da i-ésima linha e da j-ésima coluna da matriz $\mathbf{a}$. O ij-*ésimo menor* de $\mathbf{a}$ é, por definição, o determinante $\det \mathbf{M}_{ij}$ o qual se chama também o *menor relativo ao elemento* a_{ij}.

Resulta imediatamente do Teorema 19.8 que se $\mathbf{a} = [e_1, v_2, \ldots, v_n]$ é uma matriz cuja primeira coluna é $(1, 0, \ldots, 0)$ então $\det \mathbf{a} = \det \mathbf{M}_{11}$. Segue-se que se $\mathbf{a} = [e_i, v_2, \ldots, v_n]$ então

$$\det \mathbf{a} = (-1)^{i-1} \det \mathbf{M}_{i1} = (-1)^{i+1} \det \mathbf{M}_{i1}$$

pois este caso se reduz ao anterior mediante $i-1$ transposições de linhas. Portanto, dada $\mathbf{a} = [a_{ij}] = [v_1, \ldots, v_n]$, com

$$v_1 = \sum_i a_{i1} e_i,$$

tem-se

$$\det \mathbf{a} = \sum_i a_{i1} \det[e_i, v_2, \ldots, v_n],$$

logo

$$\det \mathbf{a} = \sum_i (-1)^{i+1} a_{i1} \det \mathbf{M}_{i1}\,.$$

A fórmula acima, chamada o desenvolvimento de $\det \mathbf{a}$ segundo a primeira coluna, reduz o cálculo do determinante de sua matriz $n \times n$ ao de n determinantes de matrizes $(n-1) \times (n-1)$.

Mais geralmente, podemos fixar um inteiro arbitrário j, com $1 \leq j \leq n$, e efetuar o desenvolvimento do determinante de $\mathbf{a}$ segundo a j-ésima coluna. Para isso, observamos que se

$$\mathbf{a} = [v_1, \ldots, v_{j-1}, e_i, v_{j+1}, \ldots, v_n]$$

é uma matriz cuja j-ésima coluna é e_i então $\det \mathbf{a} = (-1)^{i+j} \det \mathbf{M}_{ij}$, pois $\mathbf{a}$ pode ser transformada numa matriz cuja primeira coluna é e_1 mediante $i-1$ transposições de linha e $j-1$ transposições de coluna, o que provoca $i+j-2$ mudanças de sinal em seu determinante, e isto é o mesmo que multiplicá-lo por $(-1)^{i+j}$.

Assim, dada $\mathbf{a} = [v_1, \ldots, v_n]$, com

$$v_j = \sum_i a_{ij} e_i,$$

temos

$$\det \mathbf{a} = \sum_i a_{ij} \det[v_1, \ldots, v_{j-1}, e_i, v_{j+1}, \ldots, v_n],$$

o que fornece:

$$\det \mathbf{a} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det \mathbf{M}_{ij} .$$

Este é o *desenvolvimento do determinante de* **a** *segundo a* j*-ésima coluna*.

Como o determinante de **a** é igual ao de sua transposta $\mathbf{a}^T$, vale uma fórmula análoga, que é o *desenvolvimento segundo a* i*-ésima linha*:

$$\det \mathbf{a} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det \mathbf{M}_{ij} . \qquad (*)$$

Estas fórmulas são úteis nos cálculos, principalmente quando se escolhe uma linha ou uma coluna com diversos elementos iguais a zero. Elas também podem ser usadas como ponto de partida para uma definição indutiva do determinante de uma matriz.

Resulta da fórmula (*) que, para quaisquer $i, j = 1, \ldots, n$, temos

$$\sum_{k=1}^{n} (-1)^{j+k} a_{ik} \det \mathbf{M}_{jk} = (\det \mathbf{a}) \cdot \delta_{ij} , \qquad (**)$$

ou seja, este somatório, é igual a det **a** quando $i = j$ e igual a zero quando $i \neq j$. Com efeito, o caso $i = j$ é simplesmente a fórmula (*) e, para $i \neq j$, o somatório acima é o desenvolvimento, segundo a j-ésima linha, do determinante de uma matriz **a** na qual as linhas i e j são iguais, portanto seu determinante é zero.

O primeiro membro da igualdade (**) lembra um produto de matrizes. De fato, podemos introduzir uma matriz adj **a**, chamada a *adjunta clássica de* **a**, pondo, para $i, j = 1, \ldots, n$, $(\text{adj}\,\mathbf{a})_{ij} = (-1)^{i+j} \det \mathbf{M}_{ji}$. Então o primeiro membro da igualdade (**) é o ij-ésimo elemento do produto $\mathbf{a} \cdot \text{adj}\,\mathbf{a}$, enquanto o segundo membro é o ij-ésimo elemento da matriz (det **a**).$\mathbf{I}_n$. Portanto, a igualdade (**) significa que

$$\mathbf{a} \cdot \text{adj}\,\mathbf{a} = (\det \mathbf{a}) \cdot \mathbf{I}_n .$$

Segue-se daí que se a matriz $\mathbf{a}$ é invertível (isto é, se det $\mathbf{a} \neq 0$) então sua inversa se exprime como

$$\mathbf{a}^{-1} = \frac{1}{\det \mathbf{a}} \cdot \operatorname{adj} \mathbf{a}.$$

Esta fórmula para $\mathbf{a}^{-1}$ também pode ser obtida com auxílio da Regra de Cramer. Com efeito, o problema de obter uma matriz $\mathbf{a}^{-1} = [w_1, \ldots, w_n]$ tal que $\mathbf{a}\mathbf{a}^{-1} = \mathbf{I}_n$ pode ser considerado como n sistemas de equações lineares $\mathbf{a} \cdot w_j = e_j$ $(j = 1, \ldots, n)$. Pela Regra de Cramer, a solução $w_j = (x_{1j}, \ldots, x_{nj})$ de cada um desses sistemas é dada por

$$x_{ij} = \frac{\det \mathbf{a}[i; e_j]}{\det \mathbf{a}} = (-1)^{i+j} \det \mathbf{M}_{ji}\,,$$

pois $\mathbf{a}[i; e_j] = [v_1, \ldots, v_{i-1}, e_j, v_{i+1}, \ldots, v_n]$ se $\mathbf{a} = [v_1, \ldots, v_n]$.

Uma *submatriz* da matriz $\mathbf{a} \in M(m \times n)$ é uma matriz obtida a partir de $\mathbf{a}$ pela omissão de algumas de suas linhas e/ou colunas.

Mais precisamente, dois subconjuntos

$$R = \{i_1 < \cdots < i_r\} \subset \{1, 2, \ldots, m\}$$

e

$$S = \{j_1 < \cdots < j_s\} \subset \{1, 2, \ldots, n\}$$

determinam a submatriz $\mathbf{a}_{RS} = [a_{i_\alpha j_\beta}] \in M(r \times s)$ da matriz $\mathbf{a}$, obtida pela omissão das linhas de $\mathbf{a}$ cujos índices não pertencem a R e das colunas cujos índices não estão em S.

Teorema 19.9. *O posto de uma matriz* $\mathbf{a} \in M(m \times n)$ *é o maior número* r *tal que* $\mathbf{a}$ *possui uma submatriz* $\mathbf{a}_{RS} \in M(r \times r)$, *com* det $\mathbf{a}_{RS} \neq 0$.

Demonstração: Seja r o posto de $\mathbf{a} = [v_1, \ldots, v_n]$. Existe $S = \{j_1 < \cdots < j_r\} \subset \{1, 2, \ldots, n\}$ tal que $\{v_{j_1}, \ldots, v_{j_r}\}$ é L.I., portanto a matriz $[v_{j_1}, \ldots, v_{j_r}] \in M(m \times r)$ tem posto r. Como o posto segundo linhas é igual ao posto segundo colunas (Teorema 8.2), existe um subconjunto $R = \{i_1 < \cdots < i_r\} \subset \{1, 2, \ldots, m\}$ tal que as linhas de $[v_{j_1}, \ldots, v_{j_r}]$ cujos índices pertencem a R são L.I.. Isto significa que a matriz quadrada $\mathbf{a}_{RS} \in M(r \times r)$ é invertível, logo det $\mathbf{a}_{RS} \neq 0$. Mas é claro que o posto de uma submatriz de $\mathbf{a}$ é menor do que ou igual ao posto de $\mathbf{a}$. Portanto, nenhuma submatriz $k \times k$ de $\mathbf{a}$, com $k > r$, pode ter determinante $\neq 0$. □

Na Seção 17 (vide Observação no final de 17.G) foi provado que uma matriz simétrica é positiva se, e somente se, seus pivôs são todos positivos. Usaremos agora determinantes para provar outro critério de positividade de uma matriz, o qual é tradicional e elegante porém em dimensões mais altas é muito menos prático do que o teste dos pivôs.

Para cada $k = 1, \ldots, n$ o *menor principal* de ordem k da matriz $\mathbf{a} = [a_{ij}] \in M(n \times n)$ é o determinante da submatriz principal $\mathbf{a}_k = [a_{ij}]$, $1 \leq i, j \leq k$.

Teorema 19.10. *A fim de que uma matriz simétrica* $\mathbf{a} = [a_{ij}] \in M(n \times n)$ *seja positiva é necessário e suficiente que seus menores principais sejam todos positivos.*

Demonstração: No início de 17.G. viu-se que se $\mathbf{a}$ é positiva então suas submatrizes principais $\mathbf{a}_k$ são positivas, logo $\det \mathbf{a}_k > 0$ para $k = 1, \ldots, n$. Suponha, reciprocamente, que todos os menores principais sejam positivos. Tomando a decomposição $\mathbf{a} = \mathbf{lu}$ (veja item 5º em 17.F), resulta que $\mathbf{a}_k = \mathbf{l}_k \mathbf{u}_k$, para $k = 1, \ldots, n$, $\mathbf{l}_k$ e $\mathbf{u}_k$ indicam as submatrizes principais $k \times k$ de $\mathbf{l}$ e $\mathbf{u}$ respectivamente. Segue-se do Exemplo 19.5 que $\det \mathbf{l}_k = 1$, pois todos os elementos da diagonal de $\mathbf{l}_k$ são iguais a 1. Logo $\det \mathbf{a}_k = \det(\mathbf{l}_k \mathbf{u}_k) = \det \mathbf{u}_k = u_{11} u_{22} \ldots u_{kk}$ (produto dos elementos da diagonal de $\mathbf{u}_k$). Portanto $u_{11} = \det \mathbf{a}_1 > 0$ e $u_{kk} = \det \mathbf{a}_k / \det \mathbf{a}_{k-1} > 0$ para $k = 2, \ldots, n$. Assim todos os pivôs u_{kk} da decomposição $\mathbf{a} = \mathbf{lu}$ são positivos, logo a matriz simétrica $\mathbf{a}$ é positiva. □

Apêndice

Uma *permutação* dos inteiros $1, 2, \ldots, n$ é uma bijeção $\sigma \colon J_n \to J_n$, onde $J_n = \{1, 2, \ldots, n\}$. Um exemplo particular de permutação, quando $n > 1$, é dado pelas transposições. Uma *transposição* $\tau \colon J_n \to J_n$ é definida fixando-se dois elementos $i \neq j$ em J_n, pondo $\tau(i) = j$, $\tau(j) = i$ e $\tau(k) = k$ se $k \notin \{i, j\}$.

Se $\sigma, \sigma' \colon J_n \to J_n$ são permutações então a função composta $\sigma \circ \sigma' \colon J_n \to J_n$ também é uma permutação, chamada o *produto* das permutações σ e σ' e indicada pela notação $\sigma\sigma'$. A inversa $\sigma^{-1} \colon J_n \to J_n$ de uma permutação é ainda uma permutação, caracterizada pelo fato de que $\sigma\sigma^{-1} = \sigma^{-1}\sigma = \iota_n$ = permutação identidade $J_n \to J_n$. Se

τ é uma transposição então $\tau\tau = \iota_n$, ou seja, $\tau = \tau^{-1}$.

Às vezes se representa uma permutação $\sigma\colon J_n \to J_n$ pela lista ordenada $(\sigma(1), \ldots, \sigma(n))$ dos valores que ela assume nos números $1, 2, \ldots, n$.

O conjunto das permutações dos n números $1, 2, \ldots, n$ tem $n! = 1 \cdot 2 \ldots n$ elementos.

Toda permutação $\sigma\colon J_n \to J_n$ se escreve (de várias maneiras diferentes) como produto de transposições $\sigma = \tau_1 \cdot \tau_2 \ldots \tau_k$.

Dada a permutação σ, o número k de transposições tais que $\sigma = \tau_1 \cdot \tau_2 \ldots \tau_k$ pode variar, porém não a sua paridade. Noutras palavras, se σ puder ser expressa como produto de um número par de transposições então não poderá ser expressa como produto de um número ímpar de transposições. Ou ainda, se $\sigma = \tau_1 \cdot \tau_2 \ldots \tau_k$ onde as τ_i são transposições então $(-1)^k$ é um número que depende apenas de σ. Este número, igual a ± 1, é representado pela notação ε_σ.

Quando $\varepsilon_\sigma = +1$, diz-se que σ é uma *permutação par*. Se $\varepsilon_\sigma = -1$, diz-se que σ é uma *permutação ímpar*. Tem-se $\varepsilon_{\sigma\sigma'} = \varepsilon_\sigma \cdot \varepsilon_{\sigma'}$, ou seja, o produto de permutações pares é par, o produto de uma permutação par por uma permutação ímpar é ímpar e o produto de duas permutações ímpares é par. É claro também que $\varepsilon_\sigma = \varepsilon_{\sigma^{-1}}$.

Exercícios

19.1. Sejam $E_1, \ldots, E_r$ espaços vetoriais de dimensões $n_1, \ldots, n_r$ respectivamente. Defina função r-linear $f\colon E_1 \times \cdots \times E_r \to \mathbb{R}$ e prove que o conjunto $\mathcal{L}(E_1, \ldots, E_r; \mathbb{R})$ dessas funções é um espaço vetorial de dimensão $n_1 \cdot n_2 \cdot \ldots \cdot n_r$.

19.2. Sejam E um espaço vetorial de dimensão n e $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ uma base. Para cada seqüência crescente $J = \{j_1 < j_2 < \cdots < j_r\}$ de r inteiros de 1 até n, seja escolhido um número a_J. Prove que existe uma (única) forma r-linear alternada $f\colon E \times \cdots \times E \to \mathbb{R}$ tal que $f(u_{j_1}, \ldots, u_{j_r}) = a_J$ para toda seqüência $J = \{j_1 < \cdots < j_r\}$. Conclua que o espaço vetorial $\mathcal{A}_r(E)$ tem dimensão igual a $\binom{n}{r}$.

19.3. Dados os funcionais lineares $f_1, \ldots, f_r \colon E \to \mathbb{R}$, defina a forma r-linear alternada $f = f_1 \wedge \ldots \wedge f_r \colon E \times \cdots \times E \to \mathbb{R}$ pondo $f(v_1, \ldots, v_r) = \det(f_i(v_j))$. ("Produto exterior" dos funcionais $f_1, \ldots, f_r$.) Prove que $f_1 \wedge \ldots \wedge f_r \neq 0$ se, e somente se, $f_1, \ldots, f_r$ são L.I.. Prove também que se $\{f_1, \ldots, f_n\} \subset E^*$ é uma base então as formas $f_J = f_{j_1} \wedge \ldots \wedge f_{j_r}$, para todo $J = \{j_1 < \cdots < j_r\} \subset \{1, 2, \ldots, n\}$, constituem uma base para $\mathcal{A}_r(E)$.

19.4. Chama-se *gramiano* dos vetores $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ ao número

$$\gamma(v_1, \ldots, v_k) = \det(\langle v_i, v_j \rangle),$$

determinante da matriz de Gram $\mathbf{g}(v_1, \ldots, v_k)$. Prove:

(a) $\gamma(v_1, \ldots, v_k) > 0$ se, e somente se, os vetores $v_1, \ldots, v_k$ são linearmente independentes.

(b) Se v_1 é perpendicular a $v_2, \ldots, v_k$ então $\gamma(v_1, \ldots, v_k) = |v_1|^2 \cdot \gamma(v_2, \ldots, v_k)$.

19.5. Com a notação do exercício anterior, sejam w_1 a projeção ortogonal do vetor v_1 sobre o subespaço gerado por $v_2, \ldots, v_r$ e $h_1 = v_1 - w_1$, logo h_1 é perpendicular aos v_j com $2 \leq j \leq r$. Prove que $\gamma(v_1, \ldots, v_r) = |h_1|^2 \gamma(v_2, \ldots, v_r)$.

19.6. O *paralelepípedo* gerado pelos vetores linearmente independentes $v_1, \ldots, v_r \in \mathbb{R}^n$ é o conjunto $P[v_1, \ldots, v_r]$ das combinações lineares $t_1 v_1 + \cdots + t_r v_r$, onde $0 \leq t_i \leq 1$. O *volume* (r-dimensional) do paralelepípedo é definido por indução. Se $r = 1$, ele se reduz ao segmento de reta $[0, v_1]$, cujo "volume" uni-dimensional é, por definição, $|v_1|$. Supondo definido o volume de um paralelepípedo de dimensão $r - 1$, põe-se

$$\text{vol } P[v_1, \ldots, v_r] = |h_1| \cdot \text{vol } P[v_2, \ldots, v_r],$$

onde $|h_1|$ é a altura do paralelepípedo, isto é, $h_1 = v_1 - w_1$ e w_1 é a projeção ortogonal de v_1 sobre o subespaço gerado por $v_2, \ldots, v_r$. Prove que

$$\text{vol } P[v_1, \ldots, v_r] = \sqrt{\gamma(v_1, \ldots, v_r)} = \sqrt{\det(\langle v_i, v_j \rangle)}.$$

19.7. Seja **a** a matriz quadrada invertível cujas colunas são os vetores $v_1, \ldots, v_n \in \mathbb{R}^n$. Prove que $\gamma(v_1, \ldots, v_n) = (\det \mathbf{a})^2$ e conclua que o paralelepípedo gerado pelos vetores $v_1, \ldots, v_n$ tem volume igual a $|\det \mathbf{a}|$.

19.8. Seja $A\colon \mathbb{R}^n \to \mathbb{R}^n$ um operador linear invertível. Para todo paralelepípedo n-dimensional $X \subset \mathbb{R}^n$, prove que a imagem $A(X)$ é um paralelepípedo tal que $\operatorname{vol} A(X) = |\det A| \cdot \operatorname{vol} X$.

19.9. Calcule o determinante da matriz

$$\begin{bmatrix} 0 & 0 & 0 & a_{14} \\ 0 & 0 & a_{23} & a_{24} \\ 0 & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

e generalize o resultado para uma matriz $[a_{ij}] \in M(n \times n)$ na qual $a_{ij} = 0$ quando $i + j \leq n$.

19.10. Se a matriz triangular **b** resulta de **a** pelo processo gaussiano de eliminação, prove que $\det \mathbf{b} = (-1)^t \det \mathbf{a}$, onde t é o número de transposições de linhas feitas durante o escalonamento. (Escalonamento é o modo mais eficaz de calcular o determinante de uma matriz $n \times n$ quando $n \geq 4$.)

19.11. Escalonando a matriz de Vandermonde

$$\mathbf{v} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ x_1 & x_2 & x_3 & \ldots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \ldots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \ldots & x_n^{n-1} \end{bmatrix}$$

mostre que seu determinante é igual a

$$\prod_{i>j} (x_i - x_j),$$

logo **v** é invertível se, e somente se, os números $x_1, x_2, \ldots, x_n$ são dois a dois distintos. Como aplicação, mostre que, dados $n + 1$ pares de números $(x_o, y_o), \ldots, (x_n, y_n)$, onde $x_o < x_1 < \cdots < x_n$, existe um, e somente um, polinômio p de grau $\leq n$ tal que $p(x_o) = y_o, \ldots, p(x_n) = y_n$.

19.12. Use eliminação gaussiana (escalonamento) para calcular os determinantes das seguintes matrizes:

$$\begin{bmatrix} 1 & -2 & 1 & -1 \\ 1 & 5 & -7 & 2 \\ 3 & 1 & -5 & 3 \\ 2 & 3 & -6 & 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 2 & 1 & 3 & 2 \\ 3 & 0 & 1 & -2 \\ 1 & -1 & 4 & 3 \\ 2 & 2 & -1 & 1 \end{bmatrix}$$

19.13. Proponha e resolva um sistema linear com o uso da regra de Cramer e calcule pelo menos a inversa de uma matriz usando a adjunta clássica.

19.14. Calcule os determinantes das matrizes

$$\begin{bmatrix} 1+a & b & c \\ a & 1+b & c \\ a & b & 1+c \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} \mathbf{0} & \mathbf{I}_m \\ \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

onde os zeros representam matrizes de dimensões adequadas.

19.15. Analise o seguinte argumento segundo o qual toda matriz anti-simétrica tem determinante igual a zero: "Tem-se $\mathbf{a}^T = -\mathbf{a}$, logo $\det \mathbf{a} = \det \mathbf{a}^T = \det(-\mathbf{a}) = -\det \mathbf{a}$, logo $\det \mathbf{a} = 0$." Contraste com

$$\det \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

19.16. Defina o *produto vetorial* de n vetores $v_1, \ldots, v_n \in \mathbb{R}^{n+1}$ como o vetor $v = v_1 \times \cdots \times v_n$ tal que, para todo $w \in \mathbb{R}^{n+1}$, tem-se $\langle w, v \rangle = \det[v_1, \ldots, v_n, w]$ = determinante da matriz cujas colunas são os vetores $v_1, \ldots, v_n, w$ nesta ordem. Prove:

(a) O vetor $v = v_1 \times \cdots \times v_n \in \mathbb{R}^{n+1}$ está bem definido e é uma função n-linear alternada dos vetores $v_1, \ldots, v_n$.

(b) Seja $\mathbf{a} = [v_1, \ldots, v_n]$ a matriz $(n+1) \times n$ cujas colunas são $v_1, \ldots, v_n$. Para cada $i = 1, \ldots, n+1$, seja $\mathbf{a}_i \in M(n \times n)$ a matriz obtida de $\mathbf{a}$ pela omissão da i-ésima linha. Prove que a i-ésima coordenada do vetor $v = v_1 \times \cdots \times v_n$ é igual a $(-1)^{n+i+1} \det \mathbf{a}_i$.

(c) O produto vetorial $v = v_1 \times \cdots \times v_n$ é ortogonal a $v_1, \ldots, v_n$.

(d) Tem-se $v = v_1 \times \cdots \times v_n = 0$ se, e somente se, $v_1, \ldots, v_n$ são L.D. .

(e) Quando $v \neq 0$, a norma $|v| = |v_1 \times \cdots \times v_n|$ é igual ao volume do paralelepípedo n-dimensional $P[v_1, \ldots, v_n] \subset \mathbb{R}^{n+1}$.

[Sugestão: calcule vol $P[v, v_1, \ldots, v_n]$, levando em conta (c).)

(f) Quando os vetores $v_1, \ldots, v_n$ são L.I., tem-se $\det[v_1, \ldots, v_n, v_1 \times \cdots \times v_n] > 0$.

(g) O produto vetorial $v = v_1 \times \cdots \times v_n$ é o único vetor de $\mathbb{R}^{n+1}$ com as propriedades (c), (d), (e), (f) acima.

19.17. Para cada $i = 1, \ldots, n+1$, seja $\mathbf{a}_i \in M(n \times n)$ a matriz obtida omitindo a i-ésima linha de $\mathbf{a} \in M((n+1) \times n)$. Prove que

$$\det(\mathbf{a}^T \mathbf{a}) = \sum_{i=1}^{n+1} (\det \mathbf{a}_i)^2. \quad \text{(Identidade de Lagrange.)}$$

[Sugestão: use (e) acima e o Exercício 19.6.]

19.18. Prove que todo operador ortogonal com determinante positivo possui uma raiz quadrada ortogonal. Ela é única?

20

O Polinômio Característico

Boa parte da importância dos determinantes em Álgebra Linear se deve ao polinômio característico, o qual já tivemos ocasião de utilizar em dimensão 2. Nesta seção, já de posse da noção geral de determinante, vamos considerar esse polinômio em dimensões arbitrárias.

Seja $A\colon E \to E$ um operador linear num espaço vetorial E, de dimensão finita. A fim de que um número real λ seja autovalor de A, é necessário e suficiente que exista $v \neq 0$ em E tal que $(A - \lambda I)v = 0$, ou seja, que o operador $A - \lambda I\colon E \to E$ tenha núcleo não-trivial e portanto não seja invertível. Segundo o Teorema 19.6, isto acontece se, e somente se, $\det(A - \lambda I) = 0$.

Conforme resulta da definição clássica de determinante, $\det(A - \lambda I)$ é um polinômio de grau n em λ, cujo termo líder é $(-1)^n \lambda^n$. Ele é chamado o *polinômio característico* do operador A e é representado por $p_A(\lambda)$. Assim,

$$p_A(\lambda) = \det(A - \lambda I).$$

As raízes (reais ou complexas) da equação algébrica $p_A(\lambda) = 0$ são chamadas as *raízes características* do operador A. Do que foi dito acima, segue-se que os autovalores do operador linear A são suas raízes características reais.

Se E possui produto interno, o polinômio característico do operador adjunto $A^*\colon E \to E$ coincide com o do operador A pois

$$p_{A^*}(\lambda) = \det(A^* - \lambda I) = \det(A - \lambda I)^* = \det(A - \lambda I) = p_A(\lambda).$$

O polinômio característico de uma matriz quadrada $\mathbf{a} \in M(n \times n)$ é, por definição, $p_{\mathbf{a}}(\lambda) = \det(\mathbf{a} - \lambda \mathbf{I}_n)$, ou seja, é o polinômio característico $p_A(\lambda)$ do operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$ cuja matriz na base canônica é igual a $\mathbf{a}$. Mais geralmente, $p_{\mathbf{a}}(\lambda) = p_A(\lambda)$ para qualquer operador linear $A\colon E \to E$ cuja matriz, relativamente a uma base arbitrária de E, seja $\mathbf{a}$. (Vide Teorema 19.7.)

Se duas matrizes $\mathbf{a}$ e $\mathbf{b} = \mathbf{p}^{-1}\mathbf{a}\mathbf{p}$ são semelhantes então seus polinômios característicos são iguais. Com efeito, neste caso, $\mathbf{a}$ e $\mathbf{b}$ são matrizes do mesmo operador $A\colon \mathbb{R}^n \to \mathbb{R}^n$ relativamente a bases diferentes, logo $p_{\mathbf{a}}(\lambda) = p_A(\lambda) = p_{\mathbf{b}}(\lambda)$. Analogamente, se A e $B = P^{-1}AP$ são operadores semelhantes no espaço vetorial E, existem duas bases de E relativamente às quais A e B têm a mesma matriz $\mathbf{a}$, logo $p_A(\lambda) = p_{\mathbf{a}}(\lambda) = p_B(\lambda)$.

Exemplo 20.1. Se um dos operadores $A, B\colon E \to E$ (digamos, B) é invertível então $p_{AB}(\lambda) = p_{BA}(\lambda)$. Com efeito, neste caso, $BA = B(AB)B^{-1}$, logo BA e AB são operadores semelhantes. A igualdade entre os polinômios característicos de AB e BA prevalece, mesmo quando ambos os operadores, A e B, são não-invertíveis. Isto se prova usando um argumento de continuidade, assim: como o operador B tem no máximo um número finito de autovalores positivos, existe um número real $c > 0$ tal que $0 < \varepsilon < c \Rightarrow B - \varepsilon I$ invertível. Portanto, para todo ε positivo, menor do que c, os operadores $(B - \varepsilon I)A$ e $A(B - \varepsilon I)$ têm o mesmo polinômio característico. Como os coeficientes do polinômio característico do operador $B - \varepsilon I$ são evidentemente funções contínuas de ε, fazendo $\varepsilon \to 0$ concluímos que

$$p_{AB} = \lim_{\varepsilon \to 0} p_{A(B-\varepsilon I)} = \lim_{\varepsilon \to 0} p_{(B-\varepsilon I)A} = p_{BA}\,.$$

Exemplo 20.2. Um operador $A\colon E \to E$ diz-se *triangularizável* quando existe uma base $\mathcal{U}$ de E em relação à qual a matriz de A é triangular. Se a matriz de A na base $\mathcal{U} = \{u_1, \ldots, u_n\}$ é triangular inferior então, na base $\mathcal{U}' = \{u_n, \ldots, u_1\}$, a matriz de A é triangular superior. Isto significa que existem subespaços $F_o \subset F_1 \subset \cdots \subset F_n = E$, invariantes por A, tais que $\dim F_i = i$. Se A é triangularizável e, na

base $\mathcal{U}$, sua matriz $\mathbf{a} = [a_{ij}]$ é triangular superior então o polinômio característico de A é

$$p_A(\lambda) = \prod_{i=1}^{n}(a_{ii} - \lambda).$$

Com efeito, a matriz de $A - \lambda I$ na base $\mathcal{U}$ também é triangular superior, com os elementos $a_{ii} - \lambda$ na diagonal, logo seu determinante é igual ao produto desses números. (Exemplo 19.5.) Portanto, as raízes características de um operador triangularizável são todas reais, logo são autovalores desse operador. Em particular, são reais as raízes do polinômio característico de uma matriz simétrica (ou, o que é o mesmo, de um operador auto-adjunto num espaço com produto interno) pois toda matriz simétrica é semelhante a uma matriz diagonal, que é certamente triangular.

A noção de polinômio característico permite concluir que se a dimensão de E é um número ímpar então todo operador linear $A: E \to E$ possui pelo menos um autovalor. Com efeito, o polinômio característico $p_A(\lambda)$, sendo um polinômio real de grau ímpar, possui pelo menos uma raiz real.

Quando $\dim E = 2$, sabemos que o polinômio característico do operador $A: \mathbb{R}^2 \to \mathbb{R}^2$, cuja matriz na base canônica tem linhas (a, b) e (c, d), é igual a

$$\lambda^2 - (a + d)\lambda + ad - bc,$$

onde o coeficiente de λ é menos a soma $a + d$ dos elementos da diagonal dessa matriz e o termo constante, $ad - bc$, é seu determinante.

Em que pese a importância dos autovalores de A, é uma tarefa complicada a determinação dos coeficientes de p_A quando seu grau é elevado (e muito mais complicado ainda o cálculo de suas raízes). Um desses coeficientes é, entretanto, fácil de calcular: o termo independente de λ é igual a $p_A(0)$, logo é igual a $\det A$.

Por outro lado, se as raízes de p_A são $\lambda_1, \ldots, \lambda_n$, tem-se

$$p_A(\lambda) = (-1)^n(\lambda - \lambda_1) \cdots (\lambda - \lambda_n).$$

Pondo $\lambda = 0$ vem $\det A = p_A(0) = \lambda_1 \cdot \cdots \cdot \lambda_n$. Portanto o determinante de A é igual ao produto das suas raízes características, mesmo quando algumas delas são números complexos. (Como p_A é um polinômio real, suas raízes complexas, caso existam, vêm aos pares

conjugados, $\alpha + i\beta$ e $\alpha - i\beta$, com produto $\alpha^2 + \beta^2$, logo o produto das raízes de p_A é real.)

Outro termo de fácil determinação no polinômio $p_A(\lambda)$ é o coeficiente de λ^{n-1}. Na expressão clássica de $\det(A - \lambda I)$ em termos da matriz $\mathbf{a} = [a_{ij}]$ de A numa certa base, as parcelas que contém a potência λ^{n-1} resultam do produto $\Pi(a_{ii} - \lambda)$ dos termos da diagonal de $\mathbf{a} - \lambda \mathbf{I}_n$, logo são todas da forma $(-1)^{n-1} a_{ii} \lambda^{n-1}$. Portanto $(-1)^{n-1} \sum a_{ii}$ é o coeficiente de λ^{n-1} no polinômio $p_A(\lambda)$.

Novamente, a expressão

$$p_A(\lambda) = (-1)^n \prod_{i=1}^{n} (\lambda - \lambda_i)$$

mostra que o coeficiente de λ^{n-1} é igual a $(-1)^{n-1}$ vezes a soma das raízes do polinômio p_A.

Isto nos leva a concluir que, seja qual for a base escolhida em E, a soma Σa_{ii} dos elementos da diagonal da matriz de A nessa base é a mesma, igual à soma das raízes características de A, que é sempre um número real (mesmo que haja raízes complexas) pois $(\alpha + i\beta) + (\alpha - i\beta) = 2\alpha$.

Esta soma Σa_{ii} chama-se o *traço* do operador A e é designada com a notação tr A.

Segue-se do Exemplo 20.1 que tr AB = tr BA sejam quais forem os operadores lineares $A, B \colon E \to E$. (Isto também se vê diretamente, multiplicando as matrizes de A e B.)

Com esta notação, quando dim $E = 2$ o polinômio característico de um operador $A \colon E \to E$ se escreve $p_A(\lambda) = \lambda^2 - (\text{tr } A)\lambda + \det A$.

Vimos no Exemplo 20.2 que as raízes características de um operador triangularizável são todas números reais. Mostraremos agora que vale a recíproca.

Para isso, faremos uso do seguinte

Lema. *Seja* $F \subset E$ *um subespaço invariante pelo operador* $A \colon E \to E$. *Se* $A' \colon F \to F$ *representa a restrição de* A *ao subespaço* F, *então o polinômio* $p_{A'}$ *é um divisor de* p_A.

Demonstração: Sejam $\mathbf{a}'$ a matriz de A' numa base $\mathcal{U}' \subset F$ e $\mathbf{a}$ a matriz de A numa base $\mathcal{U} \supset \mathcal{U}'$. Então

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}' & \mathbf{b} \\ & \\ \mathbf{0} & \mathbf{c} \end{bmatrix} \quad \text{e} \quad \mathbf{a} - \lambda \mathbf{I}_n = \begin{bmatrix} \mathbf{a}' - \lambda \mathbf{I}_r & \mathbf{b} \\ & \\ \mathbf{0} & \mathbf{c} - \lambda \mathbf{I}_{n-r} \end{bmatrix},$$

onde $r = \dim F$ e $n = \dim E$. Pelo Teorema 19.8, temos

$$\begin{aligned} p_A(\lambda) &= \det(\mathbf{a} - \lambda \mathbf{I}_n) \\ &= \det(\mathbf{a}' - \lambda \mathbf{I}_r) \cdot \det(\mathbf{c} - \lambda \mathbf{I}_{n-r}) \\ &= p_{A'}(\lambda) \cdot q(\lambda), \end{aligned}$$

onde

$$q(\lambda) = \det(\mathbf{c} - \lambda \mathbf{I}_{n-r}).$$

Portanto $p_A(\lambda)$ é um múltiplo de $p_{A'}(\lambda)$. □

Teorema 20.1. *Se as raízes do polinômio característico* p_A *são todas reais então o operador* $A\colon E \to E$ *é triangularizável.*

Demonstração: O teorema é óbvio se $\dim E = 1$. Para prová-lo por indução, suponhamo-lo válido em dimensão $n-1$ e seja $\dim E = n$. Introduzamos (caso não exista ainda) um produto interno em E. Como A e A^* têm o mesmo polinômio característico, o operador $A^*\colon E \to E$ tem autovalor logo existe um subespaço $F \subset E$, de dimensão 1, invariante por A^*. O complemento ortogonal $F^\perp = F_{n-1}$ é um subespaço vetorial de dimensão $n-1$ em E, invariante por A, pois $A = (A^*)^*$. (Vide Teorema 13.3.) Pelo Lema, se $A'\colon F_{n-1} \to F_{n-1}$ é a restrição de A ao subespaço F_{n-1}, as raízes do polinômio característico $p_{A'}$ são também raízes de p_A, logo são todas reais. Pela hipótese de indução, existem subespaços $F_o \subset F_1 \subset \cdots \subset F_{n-1}$, com $\dim F_i = i$, invariantes por A', logo invariantes por A, o que prova o teorema. □

Exemplo 20.3. Um operador num espaço vetorial de dimensão 2 é triangularizável se, e somente se, possui ao menos um autovalor real. Por exemplo, uma rotação de ângulo θ, com $\theta \neq 0$ e $\theta \neq 180°$, não é triangularizável pois suas raízes características são $\cos\theta \pm i \operatorname{sen}\theta$, ambas complexas. (Vide Exemplo 14.2.) Já o operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, $A(x,y) = (7x-12y, 3x-5y)$, tem polinômio característico

$p_A(\lambda) = \lambda^2 - 2\lambda + 1$ com uma raiz real dupla $\lambda = 1$, logo A é triangularizável. Evidentemente, A não é diagonalizável pois se o fosse, como seu único autovalor é 1, seria igual ao operador identidade. Se quisermos achar uma base $\{u, v\} \subset \mathbb{R}^2$ na qual a matriz de A seja triangular superior, basta procurar um autovetor $u = (x, y)$, com $Au = u$, ou seja, basta achar uma solução não-trivial $u = (x, y)$ do sistema $7x - 12y = x$, $3x - 5y = y$. Uma dessas soluções é $u = (2, 1)$. Tomando, por exemplo, $v = (0, 1)$, a matriz de A na base $\{u, v\}$ é $\begin{bmatrix} 1 & -6 \\ 0 & 1 \end{bmatrix}$. Com efeito, $Au = u$ e $Av = -6u + v$.

Exemplo 20.4. Seja $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ a rotação de ângulo θ em torno do eixo z. Temos, para todo $(x, y, z) \in \mathbb{R}^3$:

$$A(x, y, z) = (x \cos\theta - y \operatorname{sen}\theta, x \operatorname{sen}\theta + y \cos\theta, z).$$

Para evitar casos especiais óbvios, suponhamos $0 < \theta < 180°$. Chamando de **a** a matriz de A na base canônica de $\mathbb{R}^3$, a matriz de $A - \lambda I$ é

$$\mathbf{a} - \lambda \mathbf{I}_3 = \begin{bmatrix} \cos\theta - \lambda & -\operatorname{sen}\theta & 0 \\ \operatorname{sen}\theta & \cos\theta - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{bmatrix}$$

logo $\det(A - \lambda I) = (1 - \lambda)(\lambda^2 - 2\lambda \cos\theta + 1) = p_A(\lambda)$.
Portanto o polinômio característico p_A tem uma raiz real 1 e duas raízes complexas $\cos\theta \pm i \operatorname{sen}\theta$. Assim, não existe em $\mathbb{R}^3$ uma base na qual a matriz de A seja triangular.

Examinando a demonstração do Teorema 20.1 vemos que, se o espaço E vem provido de um produto interno, ela fornece uma base ortonormal em relação à qual a matriz do operador $A\colon E \to E$ (cujo polinômio característico tem apenas raízes reais) é uma matriz triangular. Isto fornece a seguinte interpretação matricial para aquele teorema: *se o polinômio característico da matriz* $\mathbf{a} \in M(n \times n)$ *é um produto de fatores reais do primeiro grau então existe uma matriz ortogonal* $\mathbf{q} \in M(n \times n)$ *tal que* $\mathbf{t} = \mathbf{q}^T \mathbf{a} \mathbf{q}$ $(= \mathbf{q}^{-1} \mathbf{a} \mathbf{q})$ *é uma matriz triangular.*

Dados um operador linear $A\colon E \to E$ e um polinômio

$$p(\lambda) = a_o + a_1 \lambda + \cdots + a_m \lambda^m,$$

o símbolo $p(A)$ representará o operador

$$p(A) = a_o I + a_1 A + \cdots + a_m A^m,$$

obtido de $p(\lambda)$ substituindo-se λ^i por A^i, tendo o cuidado de lembrar que $A^o = I$, logo $\lambda^o = 1$ deve ser substituído por I.

Um resultado importante em Álgebra Linear é o Teorema de Cayley-Hamilton, segundo o qual, se $p = p_A$ é o polinômio característico do operador A então $p(A) = 0$. Verifiquemos a veracidade desta afirmação num caso particular.

Exemplo 20.5. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ dado por

$$A(x, y) = (ax + by, cx + dy).$$

O polinômio característico de A é

$$p_A(\lambda) = \lambda^2 - (a + d)\lambda + (ad - bc).$$

Vamos mostrar que o operador

$$B = p_A(A) = A^2 - (a + d)A + (ad - bc)I$$

é igual a zero. Para isto, basta verificar que $Be_1 = Be_2 = 0$. Ora, temos $Ae_1 = (a, c)$, $A^2 e_1 = (a^2 + bc, ac + cd)$, $e_1 = (1, 0)$, logo

$$\begin{aligned} Be_1 &= A^2 e_1 - (a + d)Ae_1 + (ad - bc)e_1 \\ &= (a^2 + bc, ac + cd) - (a^2 + ad, ac + cd) + (ad - bc, 0) \\ &= (0, 0). \end{aligned}$$

De maneira análoga se vê que $Be_2 = 0$. Portanto $B = 0$. Isto mostra que o Teorema de Cayley-Hamilton vale em dimensão 2.

Provaremos, a seguir, o Teorema de Cayley-Hamilton para operadores triangularizáveis. Na próxima seção mostraremos que, num espaço vetorial complexo, todo operador é triangularizável. Daí deduziremos a validez do teorema para qualquer operador num espaço real.

Teorema 20.2. *Se o polinômio característico* p_A *do operador* $A\colon E \to E$ *é o produto de fatores reais do primeiro grau então* $p_A(A) = 0$.

Demonstração: Pelo Teorema 20.1 existe uma base $\{u_1, \ldots, u_n\} \subset E$ relativamente à qual a matriz $\mathbf{a} = [a_{ij}]$ de A é triangular superior. Se escrevermos $F_o = \{0\}$ e F_i = subespaço vetorial de E gerado por $u_1, \ldots, u_i$, teremos $F_o \subset F_1 \subset \cdots \subset F_n = E$ e cada F_i é invariante

por A, ou seja, $A(F_i) \subset F_i$. Por hipótese, temos $p_A(\lambda) = (-1)^n(\lambda - \lambda_1)\dots(\lambda - \lambda_n)$. Escrevendo $B = p_A(A)$, resulta:

$$B = (-1)^n(A - a_{11}I)(A - a_{22}I)\cdots(A - a_{nn}I)$$

pois as raízes características de A são os elementos a_{ii} da diagonal da matriz **a**. Para cada $i = 1, \dots, n$, temos $Au_i = z + a_{ii}u_i$, onde $z \in F_{i-1}$, portanto $(A - a_{ii}I)u_i = z \in F_{i-1}$. Isto mostra que, para todo $i = 1, \dots, n$, o operador $B_i = A - a_{ii}I$ transforma F_i em F_{i-1}. Ora, temos $p_A(A) = B = B_1B_2\dots B_n$, logo

$$F_n = E \xrightarrow{B_n} \cdots \longrightarrow F_2 \xrightarrow{B_2} F_1 \xrightarrow{B_1} F_o = \{0\}$$

$$B = p_A(A)$$

$p_A(A)$ transforma E em $\{0\}$, isto é, $p_A(A) = 0$. □

Seja λ_o um autovalor do operador $A\colon E \to E$. A *multiplicidade geométrica* de λ_o é a dimensão do subespaço vetorial $F_{\lambda_o} = \{v \in E; Av = \lambda_o v\}$. A *multiplicidade algébrica* de λ_o é sua multiplicidade como raiz do polinômio característico de A, isto é, é o maior inteiro m tal que $p_A(\lambda) = (\lambda_o - \lambda)^m \cdot q(\lambda)$, onde $q(\lambda)$ é ainda um polinômio.

Obviamente, F_{λ_o} é um subespaço invariante por A. Restrito a esse subespaço, A coincide com $\lambda_o I$, ou seja, é simplesmente a multiplicação por λ_o. Portanto, o polinômio característico do operador $A'\colon F_{\lambda_o} \to F_{\lambda_o}$, restrição de A, é igual a $(\lambda_o - \lambda)^r$, onde r é a dimensão de F_{λ_o}, ou seja, a multiplicidade geométrica do autovalor λ_o. Pelo Lema que antecede o Teorema 20.1, o polinômio característico de A é um múltiplo de $(\lambda_o - \lambda)^r$, ou seja, $p_A(\lambda) = (\lambda_o - \lambda)^r\, q(\lambda)$. Isto prova o

Teorema 20.3. *A multiplicidade geométrica de um autovalor é menor do que ou igual à sua multiplicidade algébrica.*

Exemplo 20.6. No operador A do Exemplo 20.3, a multiplicidade algébrica do autovalor 1 é igual a 2 mas a sua multiplicidade geométrica é 1.

Teorema 20.4. *Se o operador* $A\colon E \to E$ *é auto-adjunto, ou ortogonal, as multiplicidades geométrica e algébrica de qualquer autovalor coincidem.*

Demonstração: Seja $F = F_{\lambda_o} = \{v \in E; Av = \lambda_o \cdot v\}$. Então F e $F^\perp$ são ambos subespaços invariantes por A. Como vimos no Lema que antecede o Teorema 20.1, se indicarmos respectivamente por $A' \colon F \to F$ e $A'' \colon F^\perp \to F^\perp$ as restrições de A a esses subespaços invariantes, teremos $p_A = p_{A'} \cdot p_{A''}$, ou seja, $p_A(\lambda) = (\lambda_o - \lambda)^r \cdot p_{A''}(\lambda)$, onde $r = \dim F$. Mas, pela definição de F, não pode existir em $F^\perp$ (nem em lugar algum fora de F) um autovetor correspondente ao autovalor λ_o. Logo λ_o não é raiz de $p_{A''}$. Assim, r é o maior inteiro tal que $(\lambda_o - \lambda)^r$ divide p_A, ou seja, é a multiplicidade algébrica de λ_o. □

Exercícios

20.1. Assinale V(erdadeiro) ou F(also):

( ) Os operadores A e A^* têm os mesmos autovetores.

( ) Sejam **a** a matriz do operador $A \colon \mathbb{R}^n \to \mathbb{R}^n$ na base canônica e **p** uma matriz cujas colunas são autovetores L.I. de A. Então $\mathbf{p}^{-1}\mathbf{ap}$ é diagonal.

( ) Se λ é autovalor do operador invertível A então λ^{-1} é autovalor de A^{-1}.

( ) O polinômio característico do operador $A + B$ é a soma dos polinômios característicos de A e B.

( ) Se v é um autovetor comum aos operadores A e B então v é autovetor de $A + B$ e de BA.

( ) Duas matrizes triangulares semelhantes são iguais.

20.2. Determine os polinômios característicos dos seguintes operadores:

(a) Um múltiplo αI da identidade;

(b) Uma projeção P;

(c) Uma involução;

(d) O operador de derivação $D \colon \mathcal{P}_n \to \mathcal{P}_n$.

20.3. No espaço vetorial E, de dimensão finita, com produto interno, seja $Av = \langle v, a \rangle\, b$ um operador de posto 1. Escreva a matriz de A numa base ortonormal que comece com $u_1 = a/|a|$ e, a partir daí, determine o polinômio característico de A, seus autovetores e seus autovalores com as respectivas multiplicidades.

20.4. Qual o coeficiente de λ^2 no polinômio característico $p_{\mathbf{a}}(\lambda)$ de uma matriz $\mathbf{a} = [a_{ij}] \in M(3 \times 3)$?

20.5. Seja $F_1 \subset E$ um subespaço invariante pelo operador linear $A \colon E \to E$. Se $E = F_1 \oplus F_2$, seja P a projeção sobre F_2 paralelamente a F_1. Indique com $A_1 \colon F_1 \to F_1$ a restrição de A a F_1 e com $A_2 \colon F_2 \to F_2$ o operador dado por $A_2 v = PAv$, $v \in F_2$. Prove que $p_A(\lambda) = p_{A_1}(\lambda) \cdot p_{A_2}(\lambda)$. Seja $\mathcal{U}$ uma base de E, formada por uma base de F_1 seguida por outra de F_2. Mostre que a matriz de A na base $\mathcal{U}$ tem a forma

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{b} \\ 0 & \mathbf{a}_2 \end{bmatrix}$$

onde $\mathbf{a}_1$ e $\mathbf{a}_2$ são as matrizes de A_1 e A_2 respectivamente. Dê um enunciado mais simples, em termos de matrizes, para a proposição cuja tese é $p_A(\lambda) = p_{A_1}(\lambda) \cdot p_{A_2}(\lambda)$.

20.6. Determine o polinômio característico, ache os autovalores e exiba uma base de autovetores para a matriz

$$\begin{bmatrix} 4 & -3 & 1 & 1 \\ 2 & -1 & 1 & 1 \\ 0 & 0 & -4 & 3 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

20.7. Determine o polinômio característico e os autovalores da matriz

$$\begin{bmatrix} 4 & -3 & a_3 & a_4 & a_5 & a_6 \\ 2 & -1 & b_3 & b_4 & b_5 & b_6 \\ 0 & 0 & -4 & 3 & c_5 & c_6 \\ 0 & 0 & 2 & 1 & d_5 & d_6 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 3 & -1 \end{bmatrix}$$

20.8. Quais são os autovetores do operador de derivação $D \colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$?

20.9. Assinale V(erdadeiro) ou F(also):

( ) Se um operador é diagonalizável, todas as suas matrizes triangulares são diagonais.

( ) Seja **a** uma matriz triangular não-diagonal. Se todos os elementos da diagonal de **a** forem iguais, **a** não é diagonalizável.

( ) Uma matriz 3×3 que tem dois autovalores distintos é triangularizável.

20.10. Sejam $A, B \colon E \to E$ operadores cujas raízes características são todas reais. Se $AB = BA$, prove que existe uma base na qual as matrizes de A e B são ambas triangulares.

20.11. Ache uma base de $\mathbb{R}^3$ na qual o operador $A(x, y, z) = (x + 2y + 3z, 4y + 6z, -y - z)$ tem uma matriz triangular. Exiba essa matriz.

20.12. Determine o polinômio caracterítico da matriz

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ a_{n-1} & a_{n-2} & a_{n-3} & \dots & a_o \end{bmatrix}$$

20.13. Obtenha o polinômio característico e os autovalores (com as respectivas multiplicidades, algébricas e geométricas) do operador $A \colon \mathbb{R}^n \to \mathbb{R}^n$ cuja matriz na base canônica tem todos os elementos iguais a 1.

[Sugestão: use o Exercício 20.3.]

20.14. Prove que o módulo do determinante de um operador invertível é igual ao produto dos seus valores singulares.

20.15. Sejam $A, B \colon E \to E$ operadores lineares não-negativos e X a raiz quadrada não-negativa de A. Prove:

(a) AB e XBX têm o mesmo polinômio característico, logo $\det(I + AB) = \det(I + XBX)$.

(b) O operador XBX é não-negativo e

$$\det(I + XBX) = \prod_{i=1}^{n} (1 + \lambda_i),$$

onde $\lambda_1, \ldots, \lambda_n$ são os autovalores de XBX.

(c) Conclua que $\det(I+AB) \geq 1$. Em particular, $I+AB$ é invertível.

21

Espaços Vetoriais Complexos

Embora a noção de espaço vetorial tenha sentido (e interesse) sobre um corpo qualquer, neste livro os vetores (pelo menos até agora) vêm sendo multiplicados apenas por números reais. Esta opção foi feita por uma série de razões, das quais destacaremos duas. Em primeiro lugar, ao nos limitarmos aos números reais, não temos que nos preocupar com as peculiaridades dos vários corpos possíveis o que, num livro introdutório, traria o risco de focalizar a atenção no acidental. Assim ficou mais fácil nos concentrarmos em questões realmente essenciais, sem maior perda de tempo. Em segundo lugar, porque o caso real é, sem dúvida, o mais importante. Entretanto, o corpo dos números reais não é algebricamente completo: nem todo polinômio com coeficientes reais possui raiz real. O corpo dos números complexos não sofre dessa deficiência. Isto torna necessário que alguns teoremas referentes a espaços vetoriais reais utilizem números complexos em sua demonstração (como foi feito, um tanto disfarçadamente, no Teorema 12.1). Na presente seção, é introduzido o conceito de espaço vetorial complexo e é mostrado explicitamente como ele pode ser útil para demonstrar teoremas sobre espaços vetoriais reais.

Nesta seção, os espaços vetoriais que viemos estudando até agora serão chamados *espaços vetoriais reais* e as transformações lineares

neles definidas serão chamadas $\mathbb{R}$-*lineares*. Analogamente, as matrizes até agora consideradas chamar-se-ão *matrizes reais*. A razão para essa qualificação é que introduziremos aqui os espaços vetoriais complexos.

Um *espaço vetorial complexo* é um conjunto E, cujos elementos são chamados *vetores*, no qual estão definidas duas operações: a *adição*, que faz corresponder a cada par de vetores $u, v \in E$ um vetor $u + v$, chamado a *soma* de u e v, e a *multiplicação por um número complexo*, que a cada número complexo ζ e a cada vetor $v \in E$ faz corresponder um vetor $\zeta \cdot v = \zeta v$, chamado o *produto* de ζ por v. Essas operações devem cumprir as mesmas condições impostas na Seção 1 para os espaços vetoriais reais. Em particular, se $v \in E$ e $\zeta = \alpha + i\beta$ então $\zeta v = \alpha v + i\beta v = \alpha v + \beta(iv)$.

Exemplo 21.1. O conjunto $\mathbb{C}^n$ de todas as listas $u = (\xi_1, \ldots, \xi_n)$, $v = (\zeta_1, \ldots, \zeta_n)$ de n números complexos, com as definições $u + v = (\xi_1 + \zeta_1, \ldots, \xi_n + \zeta_n)$, $\zeta \cdot u = (\zeta \cdot \xi_1, \ldots, \zeta \cdot \xi_n)$, é um espaço vetorial complexo. Também o conjunto $\mathcal{F}(X; \mathbb{C})$ de todas as funções $f \colon X \to \mathbb{C}$, definidas num conjunto arbitrário X, com valores complexos, é um espaço vetorial complexo quando munido das definições óbvias para $f + g$ e $\zeta \cdot f$. O conjunto $M(m \times n; \mathbb{C})$ das matrizes complexas $m \times n$ também é um espaço vetorial complexo.

As definições de dependência linear, geradores, subespaço, base, dimensão, etc. se fazem para os espaços vetoriais complexos da mesma maneira como foram feitas para os reais. Na realidade, tudo o que foi dito e demonstrado nas Seções 1 a 9 vale para espaços vetoriais complexos e as transformações lineares entre eles, as quais chamaremos $\mathbb{C}$-*lineares*.

Às vezes, uma base do espaço vetorial complexo E será chamada uma $\mathbb{C}$-*base*.

Um espaço vetorial complexo E pode, de modo natural, ser considerado como um espaço vetorial real: basta que se considere apenas a multiplicação dos vetores de E por números reais. Analogamente, toda transformação $\mathbb{C}$-linear $A \colon E \to F$ entre espaços vetoriais complexos é, a fortiori, $\mathbb{R}$-linear. Quando for conveniente, usaremos a notação $A_r \colon E \to F$ para indicar essa transformação $\mathbb{R}$-linear, que se chama a *descomplexificada* de A.

Exemplo 21.2. O conjunto $\mathbb{C}$ dos números complexos é um espaço vetorial complexo de dimensão 1, logo todo número complexo

$\alpha + i\beta \neq 0$ fornece uma base para $\mathbb{C}$. Em particular, $\{1\}$ é uma $\mathbb{C}$-base. Considerando $\mathbb{C}$ como espaço vetorial real, o conjunto $\{1, i\} \subset \mathbb{C}$ é L.I. pois $\alpha \cdot 1 + \beta \cdot i = 0$, com $\alpha, \beta \in \mathbb{R}$ implica $\alpha = \beta = 0$. Na realidade, $\{1, i\}$ é uma $\mathbb{R}$-base pois todo número complexo $\alpha + i\beta = \alpha \cdot 1 + \beta \cdot i$ é uma combinação linear real de 1 e i. Em virtude da unidimensionalidade, os operadores $\mathbb{C}$-lineares $A \colon \mathbb{C} \to \mathbb{C}$ consistem simplesmente na multiplicação por um número complexo fixado $\alpha + i\beta$. Assim, para todo $v = x + iy \in \mathbb{C}$, tem-se $Av = (\alpha + i\beta)v = (\alpha + i\beta)(x + iy) = \alpha x - \beta y + i(\beta x + \alpha y)$.

A correspondência $x + iy \leftrightarrow (x, y)$ é um isomorfismo natural entre os espaços vetoriais reais $\mathbb{C}$ e $\mathbb{R}^2$. Pelo que acabamos de ver, esse isomorfismo faz corresponder a cada operador $\mathbb{C}$-linear $A \colon \mathbb{C} \to \mathbb{C}$, um operador $\mathbb{R}$-linear $A_r \colon \mathbb{R}^2 \to \mathbb{R}^2$, dado por

$$A_r(x, y) = (\alpha x - \beta y, \beta x + \alpha y),$$

cuja matriz na base canônica é

$$\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}.$$

Excetuemos o caso trivial $A = 0$. Escrevamos

$$\rho = \sqrt{\alpha^2 + \beta^2}$$

e tomemos $\theta \in \mathbb{R}$ tal que $\cos\theta = \alpha/\rho$, $\operatorname{sen}\theta = \beta/\rho$. Então a matriz de A_r tem a forma

$$\rho \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Isto mostra que o operador A_r é uma *semelhança*, composta da rotação de ângulo θ com a homotetia de razão $\rho > 0$. As semelhanças são, portanto, os operadores de $\mathbb{R}^2$ que correspondem aos operadores $\mathbb{C}$-lineares não-nulos $A \colon \mathbb{C} \to \mathbb{C}$.

Guiados pelo Exemplo 21.2, observamos que se $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ for uma base do espaço vetorial complexo E então o conjunto $\mathcal{U}' = \{u_1, \ldots, u_n, iu_1, \ldots, iu_n\} \subset E$ é uma $\mathbb{R}$-base, ou seja, é L.I. sobre os reais e, além disso, todo vetor $v \in E$ é uma combinação linear dos u_j e dos iu_j $(j = 1, \ldots, n)$ com coeficientes reais.

Com efeito, se $\alpha_1, \ldots, \alpha_n$, $\beta_1, \ldots, \beta_n \in \mathbb{R}$ então

$$\alpha_1 u_1 + \cdots + \alpha_n u_n + \beta_1 iu_1 + \cdots + \beta_n iu_n = 0$$

implica

$$(\alpha_1 + i\beta_1)u_1 + \cdots + (\alpha_n + i\beta_n)u_n = 0,$$

logo $\alpha_1 + i\beta_1 = \cdots = \alpha_n + i\beta_n = 0$ pois $\mathcal{U}$ é L.I. sobre os complexos. Daí resulta que $\alpha_1 = \cdots = \alpha_n = \beta_1 = \cdots = \beta_n = 0$, portanto $\mathcal{U}'$ é L.I. sobre os reais. Além disso, dado qualquer $v \in E$, existem números complexos $\alpha_1 + i\beta_1, \ldots, \alpha_n + i\beta_n$ tais que

$$v = \sum_{j=1}^{n} (\alpha_j + i\beta_j)u_j = \sum_{j=1}^{n} \alpha_j u_j + \sum_{j=1}^{n} \beta_j (iu_j).$$

Assim, $\mathcal{U}'$ é uma $\mathbb{R}$-base para E.

Com uma notação de significado evidente, podemos portanto concluir que se $\dim_{\mathbb{C}} E = n$ então $\dim_{\mathbb{R}} E = 2n$.

Seja $[a_{kj} + ib_{kj}] \in M(m \times n; \mathbb{C})$ a matriz da transformação $\mathbb{C}$-linear $A \colon E \to F$ relativamente às bases $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ e $\mathcal{V} = \{v_1, \ldots, v_m\} \subset F$. Considerando E e F como espaços vetoriais reais, A pode ser vista como uma transformação $\mathbb{R}$-linear $A_r \colon E \to F$, a *descomplexificada* de A. Relativamente às bases

$$\mathcal{U}' = \{u_1, \ldots, u_n, iu_1, \ldots, iu_n\} \subset E$$

e

$$\mathcal{V}' = \{v_1, \ldots, v_m, iv_1, \ldots, iv_m\} \subset F,$$

vejamos qual é a matriz de A_r. Temos, para $j = 1, \ldots, n$:

$$A_r u_j = Au_j = \sum_{k=1}^{m} (a_{kj} + ib_{kj})v_k = \sum_{k=1}^{m} a_{kj} v_k + \sum_{k=1}^{m} b_{kj}(iv_k),$$

$$A_r(iu_j) = A(iu_j) = i \cdot Au_j = \sum_{k=1}^{m} (-b_{kj})v_k + \sum_{k=1}^{m} a_{kj}(iv_k),$$

Portanto a matriz procurada é

$$\mathbf{c} = \begin{bmatrix} \mathbf{a} & -\mathbf{b} \\ \mathbf{b} & \mathbf{a} \end{bmatrix} \in M(2m \times 2n),$$

onde $\mathbf{a} = [a_{kj}] \in M(m \times n)$ e $\mathbf{b} = [b_{kj}] \in M(m \times n)$. Esta matriz $2m \times 2n$ chama-se a *descomplexificada* da matriz $[a_{kj} + ib_{kj}] \in M(m \times n; \mathbb{C})$. Mostraremos logo mais que, quando $m = n$, tem-se $\det \mathbf{c} \geq 0$, e que $\det \mathbf{c} = 0$ se, e somente se, $\det A = 0$.

Todo espaço vetorial real E de dimensão par $2n$ pode (de infinitas maneiras) ser considerado como espaço vetorial complexo de dimensão n de tal forma que a nova multiplicação de um número complexo por um vetor coincida com a multiplicação anterior quando esse número complexo é real. Para isso, basta considerar um operador $\mathbb{R}$-linear $J \colon E \to E$ tal que $J^2 = -I$ e definir, para cada número complexo $\zeta = \alpha + i\beta$ e cada vetor $v \in E$, o produto $\zeta \cdot v$ como $\zeta \cdot v = \alpha v + \beta J v$. (A adição de vetores continua a mesma.)

A verificação de que esta definição atende às exigências sobre as regras operacionais é imediata. Resta apenas mostrar como se acha um tal operador J. Para isso, fixamos uma base de E, a qual numeramos da forma $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$. Existe um único operador $\mathbb{R}$-linear $J \colon E \to E$ tal que $Ju_1 = v_1, \ldots, Ju_n = v_n$, $Jv_1 = -u_1, \ldots, Jv_n = -u_n$. Nesta concepção de E como espaço vetorial complexo, o operador J é simplesmente a multiplicação pelo número i, logo $v_1 = iu_1, \ldots, v_n = iu_n$ e $\{u_1, \ldots, u_n\} \subset E$ é uma $\mathbb{C}$-base.

A noção de espaço vetorial complexo foi introduzida aqui a fim de servir como instrumento para obter resultados referentes a espaços vetoriais reais.

A esse propósito, a principal vantagem dos espaços vetoriais complexos sobre os reais é a seguinte: todo operador linear $A \colon E \to E$ num espaço vetorial complexo de dimensão finita possui pelo menos um autovalor (complexo). Antes, porém, de estabelecermos e explorarmos este fato, vamos retomar uma afirmação anterior, segundo a qual tudo o que foi feito até a Seção 9 sobre espaços vetoriais reais vale igualmente para complexos. Por que não foi incluída a Seção 10?

É que se faz necessário modificar o conceito de produto interno quando se trata de um espaço vetorial complexo E. Se o produto interno for bilinear então $\langle iv, iv \rangle = i^2 \langle v, v \rangle = -\langle v, v \rangle$, logo não pode ser positivo.

O impasse é resolvido mediante a noção de *produto interno hermitiano*. Este é, por definição, uma função $E \times E \to \mathbb{C}$ que associa a cada par ordenado de vetores u, v no espaço vetorial complexo E um número complexo, representado pela notação $\langle u, v \rangle$, de tal modo que sejam cumpridas as condições seguintes, para quaisquer $u, v, u' \in E$, $\zeta \in \mathbb{C}$, onde uma barra sobre um número complexo $\zeta = \alpha + i\beta$ significa seu conjugado $\overline{\zeta} = \alpha - i\beta$:

1. $\langle u, v \rangle = \overline{\langle v, u \rangle}$;

2. $\langle u + u', v \rangle = \langle u, v \rangle + \langle u', v \rangle$;

3. $\langle \zeta u, v \rangle = \zeta \langle u, v \rangle$;

4. $\langle u, u \rangle > 0$, se $u \neq 0$.

Das propriedades 1. e 2. segue-se que $\langle u, v + v' \rangle = \langle u, v \rangle + \langle u, v' \rangle$. Com efeito,

$$\begin{aligned} \langle u, v + v' \rangle &= \overline{\langle v + v', u \rangle} = \overline{\langle v, u \rangle + \langle v', u \rangle} = \overline{\langle v, u \rangle} + \overline{\langle v', u \rangle} \\ &= \langle u, v \rangle + \langle u, v' \rangle \,. \end{aligned}$$

Analogamente se mostra que 1. e 3. implicam que $\langle u, \zeta v \rangle = \overline{\zeta} \langle u, v \rangle$. Assim, o produto interno hermitiano é *sesqui-linear*, ou seja, linear na primeira variável e anti-linear na segunda. Segue-se de 1. que $\langle u, v \rangle = \langle v, u \rangle \Leftrightarrow \langle u, v \rangle \in \mathbb{R}$.

Exemplo 21.3. No espaço $\mathbb{C}^n$, o produto interno canônico é definido, para $u = (\xi_1, \ldots, \xi_n)$ e $v = (\zeta_1, \ldots, \zeta_n)$, como

$$\langle u, v \rangle = \xi_1 \overline{\zeta}_1 + \cdots + \xi_n \overline{\zeta}_n .$$

As 4 propriedades acima são imediatamente verificadas de modo que se trata de um produto interno hermitiano.

Exemplo 21.4. Seja $E = C^0([a, b]; \mathbb{C})$ o espaço vetorial complexo formado pelas funções contínuas $f \colon [a, b] \to \mathbb{C}$, definidas no intervalo $[a, b]$ e tomando valores complexos. Um produto interno hermitiano em E pode ser definido pondo

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} \, dx,$$

para $f, g \in E$ quaisquer.

Exemplo 21.5. Em todo espaço vetorial complexo E, de dimensão finita, pode-se introduzir um produto interno hermitiano. (Na realidade, uma infinidade deles.) Basta tomar uma base $\{u_1, \ldots, u_n\} \subset E$ e, para dois vetores

$$u = \sum \xi_k u_k, \quad v = \sum \zeta_k u_k$$

quaisquer em E, pôr
$$\langle u, v\rangle = \sum \xi_k \overline{\zeta}_k.$$
Isto mostra que, como no caso real, a existência de um produto interno hermitiano num espaço vetorial complexo de dimensão finita não é uma propriedade adicional desse espaço mas apenas uma escolha que se fez dentre infinitas possibilidades.

A partir da definição de produto interno hermitiano há poucas adaptações a fazer a fim de que as restantes seções, de 10 a 20, tenham seus resultados validados (e alguns fortalecidos, como veremos logo mais).

Uma das modificações a fazer como conseqüência da sesqui-linearidade do produto interno hermitiano diz respeito ao Teorema 11.1, que passa a ter o seguinte enunciado:

Teorema 21.1. *Seja E um espaço vetorial complexo de dimensão finita, munido de um produto interno hermitiano. A correspondência que associa a cada vetor $v \in E$ o funcional linear $\phi(v) = v^*\colon E \to \mathbb{C}$, tal que $v^*(w) = \langle w, v\rangle$ para todo $w \in E$, é uma bijeção $\phi\colon E \to E^*$ tal que $(u+v)^* = u^* + v^*$ e $(\zeta v)^* = \overline{\zeta} \cdot v^*$ para quaisquer $u, v \in E$ e $\zeta \in \mathbb{C}$.*

Acima, E^* é, como antes, o *dual* de E, ou seja, o espaço vetorial complexo cujos elementos são os funcionais $\mathbb{C}$-lineares $f\colon E \to \mathbb{C}$.

A diferença entre os Teoremas 11.1 e 21.1 é que, neste último, a correspondência $v \mapsto v^*$ não é um isomorfismo entre E e E^*, pois falha a condição $(\zeta v)^* = \zeta v^*$, em vez da qual se tem apenas $(\zeta v)^* = \overline{\zeta} v^*$. Com efeito, para todo $w \in E$, tem-se
$$(\zeta v)^*(w) = \langle w, \zeta v\rangle = \overline{\zeta}\,\langle w, v\rangle = \overline{\zeta} v^*(w), \text{ portanto } (\zeta v)^* = \overline{\zeta} v^*.$$
Por isso, ela se chama um *anti-isomorfismo*.

O Teorema 21.1 pode então ser refraseado dizendo-se que a bijeção $\phi\colon E \to E^*$ nele definida é um anti-isomorfismo.

Isto não impede de maneira nenhuma que se defina a adjunta $A^*\colon F \to E$ de uma transformação $\mathbb{C}$-linear $A\colon E \to F$, entre espaços vetoriais complexos munidos de produto interno hermitiano, como a única transformação $\mathbb{C}$-linear $A^*\colon F \to E$ tal que
$$\langle Av, w\rangle = \langle v, A^*w\rangle$$
para $v \in E$, $w \in F$ quaisquer. Valem as propriedades $(A + B)^* = A^* + B^*$, $(BA)^* = A^*B^*$, $I^* = I$, $(A^*)^* = A$ e $(A^*)^{-1} = (A^{-1})^*$, nesta

última entendendo-se (como antes) que A^* é invertível se, e somente se, A é. A única diferença diz respeito à adjunta de ζA. Tem-se $(\zeta A)^* = \overline{\zeta} A^*$ quando ζ é um número complexo.

Outra mudança ocorre no Teorema 11.2, que passa a ter o enunciado abaixo. E e F são espaços vetoriais complexos, com produto interno hermitiano.

Teorema 21.2. *Se a matriz da transformação $\mathbb{C}$-linear $A: E \to F$ relativamente a bases ortonormais $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ e $\mathcal{V} = \{v_1, \ldots, v_m\} \subset F$ é $\mathbf{a} = [a_{kj}] \in M(m \times n)$ então a matriz da transformação adjunta $A^*: F \to E$ relativamente às bases V, $\mathcal{U}$ é a matriz $\mathbf{a}^* = \overline{\mathbf{a}}^T$, transposta da conjugada de $\mathbf{a}$.*

A *matriz conjugada* de $\mathbf{a} = [a_{kj}] \in M(m \times n; \mathbb{C})$ é a matriz $\overline{\mathbf{a}} = [\overline{a}_{kj}] \in M(m \times n; \mathbb{C})$ onde cada elemento $\overline{a}_{kj}$ é o número complexo conjugado do elemento correspondente a_{kj} da matriz $\mathbf{a}$.

Um operador $\mathbb{C}$-linear $A: E \to E$ chama-se *hermitiano* quando $A = A^*$, ou seja, quando $\langle Au, v\rangle = \langle u, Av\rangle$ para quaisquer $u, v \in E$. Em particular, quando $u = v$, temos $\langle Av, v\rangle = \langle v, Av\rangle$. Portanto, quando A é hermitiano, a forma quadrática $\varphi(v) = \langle v, Av\rangle$ só assume valores reais.

Uma matriz $\mathbf{a} = [a_{kj}] \in M(n \times n; \mathbb{C})$ chama-se *hermitiana* quando $\mathbf{a} = \mathbf{a}^*$, isto é, quando $a_{jk} = \overline{a}_{kj}$ para $k, j = 1, \ldots, n$. Em particular, $a_{jj} = \overline{a}_{jj}$ para todo $j = 1, \ldots, n$ portanto a diagonal de uma matriz hermitiana só possui números reais. Para matrizes reais, hermitiana é o mesmo que simétrica.

Um operador $\mathbb{C}$-linear $A: E \to E$ é hermitiano se, e somente se, sua matriz relativamente a uma (e portanto a qualquer) base ortonormal de E é uma matriz hermitiana.

Um operador $\mathbb{C}$-linear $U: E \to E$ chama-se *unitário* quando $U^* = U^{-1}$. Isto equivale a dizer que $\langle Uv, Uw\rangle = \langle v, w\rangle$ para quaisquer $v, w \in E$.

Se $U: E \to E$ é um operador unitário então, para todo $v \in E$ tem-se $|Uv| = |v|$. Vale também a recíproca. Para prová-la, usa-se a *identidade de polarização*

$$\langle u, v\rangle = \frac{1}{4}[|u+v|^2 - |u-v|^2 + i|u+iv|^2 - i|u-iv|^2],$$

cuja verificação é imediata. Sua forma é mais complicada do que a análoga real, devido ao fato de que o produto interno hermitiano não

é simétrico.

Uma matriz $\mathbf{u} \in M(n \times n; \mathbb{C})$ chama-se *unitária* quando $\mathbf{u}^* = \mathbf{u}^{-1}$. Para que isto aconteça basta que $\mathbf{u}^*\mathbf{u} = \mathbf{I}_n$, ou que $\mathbf{u}\mathbf{u}^* = \mathbf{I}_n$. A primeira destas igualdades significa que as colunas de $\mathbf{u}$ formam uma base ortonormal de $\mathbb{C}^n$ (relativamente ao produto interno hermitiano). A segunda assegura a mesma propriedade para as linhas de $\mathbf{u}$. As matrizes unitárias reais são as ortogonais.

Feitas essas observações de caráter geral, passemos a tirar proveito da estrutura complexa.

O determinante e o polinômio característico, como não têm nada a ver com produto interno, se definem de modo análogo ao caso real.

Segundo o Teorema Fundamental da Álgebra, todo polinômio

$$p(\lambda) = a_o + a_1\lambda + \cdots + a_n\lambda^n$$

com coeficientes complexos $a_o, \ldots, a_n$ (em particular, com coeficientes reais), com $a_n = (-1)^n$, se decompõe como produto

$$p(\lambda) = \prod_{k=1}^{n} (\lambda_k - \lambda)$$

de fatores do primeiro grau. Os números complexos $\lambda_1, \ldots, \lambda_n$, não necessariamente distintos, são as raízes do polinômio p, cada um deles comparecendo na lista $\lambda_1, \ldots, \lambda_n$ um número de vezes chamado sua *multiplicidade* como raiz do polinômio. A multiplicidade de λ_k é m se, e somente se, m é o maior inteiro tal que $p(\lambda)$ é divisível por $(\lambda_k - \lambda)^m$.

Seja $p_A(\lambda) = \det(A - \lambda I)$ o polinômio característico do operador $\mathbb{C}$-linear $A\colon E \to E$. O número complexo λ_o é uma raiz de p_A se, e somente se, o operador $A - \lambda_o I$ é não-invertível, ou seja, existe $v \neq 0$ em E tal que $Av = \lambda_o v$. Portanto, para espaços vetoriais complexos, as raízes características de um operador $A\colon E \to E$ coincidem com os autovalores desse operador. Como as primeiras sempre existem, segue-se que todo operador $\mathbb{C}$-linear possui autovalores (que podem ser reais ou complexos).

A existência de autovetores nos espaços vetoriais complexos implica a seguinte versão fortalecida do Teorema 20.1, da qual decorrerão todas as conclusões a que chegaremos nesta seção.

Teorema 21.3. *Todo operador* $\mathbb{C}$*-linear* $A: E \to E$ *é triangularizável.*

A demonstração é a mesma do Teorema 20.1. Só que agora não é necessário fazer hipótese adicional sobre A porque todo operador linear complexo tem autovetor.

Como anteriormente, vale a importante observação de que se E possui produto interno hermitiano, o enunciado acima pode ser tornado mais preciso: existe uma base ortonormal $\mathcal{U} \subset E$ na qual a matriz de A é triangular superior (ou inferior, se assim o quisermos).

A versão matricial desse teorema é: *para toda matriz complexa* $\mathbf{a}$ *existe uma matriz unitária* $\mathbf{u}$ *tal que* $\mathbf{u}^*\mathbf{a}\mathbf{u} = \mathbf{u}^{-1}\mathbf{a}\mathbf{u} = \mathbf{t}$ *é uma matriz triangular.*

Vejamos, a seguir, algumas conseqüências do Teorema 21.3. Na primeira delas, temos um operador linear $A: E \to E$ num espaço vetorial complexo E.

Teorema de Cayley-Hamilton. *Se* p_A *é o polinômio característico do operador* $\mathbb{C}$*-linear* $A: E \to E$ *então* $p_A(A) = 0$*.*

Demonstração: segue exatamente a linha da demonstração do Teorema 20.2 pois, em virtude do Teorema 21.3, todo operador $\mathbb{C}$-linear é triangularizável. □

Evidentemente, vale uma versão matricial de Cayley-Hamilton. Para toda matriz $\mathbf{a} \in M(n \times n; \mathbb{C})$, o polinômio característico $p_{\mathbf{a}}(\lambda)$ é exatamente o polinômio $p_A(\lambda)$, onde $A: \mathbb{C}^n \to \mathbb{C}^n$ é o operador $\mathbb{C}$-linear que tem matriz $\mathbf{a}$ na base canônica. Tem-se também $p_{\mathbf{a}} = p_A$ para qualquer operador $\mathbb{C}$-linear $A: E \to E$ cuja matriz, relativamente a uma base arbitrária em E, seja igual a $\mathbf{a}$. Qualquer que seja o polinômio $q(\lambda)$, vemos que $q(\mathbf{a}) \in M(n \times n; \mathbb{C})$ é a matriz do operador $q(A)$ na mesma base relativamente à qual $\mathbf{a}$ é a matriz de A. Portanto, *se* $p_{\mathbf{a}}$ *é o polinômio característico da matriz* $\mathbf{a} \in M(n \times n; \mathbb{C})$*, tem-se* $p_{\mathbf{a}}(\mathbf{a}) = 0$*.*

Se acontecer de a matriz $\mathbf{a}$ ser real, seu polinômio característico $p_{\mathbf{a}}$ é um polinômio real e ainda assim se tem $p_{\mathbf{a}}(\mathbf{a}) = 0$, pois todo número real é complexo. Segue-se daí o

Teorema de Cayley-Hamilton para operadores reais. *Seja* $A: E \to E$ *um operador linear num espaço vetorial real* E*. Se* p_A *é seu polinômio característico, tem-se* $p_A(A) = 0$*.*

Demonstração: Seja $\mathbf{a} \in M(n \times n)$ a matriz de A relativamente a uma certa base de E. Então $p_{\mathbf{a}}(\mathbf{a}) = 0$. Como $p_{\mathbf{a}}(\mathbf{a})$ é a matriz do operador $p_A(A)$ nessa mesma base, segue-se que $p_A(A) = 0$. □

Continuemos apresentando conseqüências do Teorema 21.3.

Sejam $\mathbf{a} = [a_{kj}]$, $\mathbf{b} = [b_{kj}]$ matrizes $n \times n$ triangulares superiores, de modo que $a_{kj} = b_{kj} = 0$ se $k > j$. O j-ésimo elemento da diagonal do produto $\mathbf{ab}$ é $(\mathbf{ab})_{jj} = \sum_r a_{jr}b_{rj} = a_{jj}b_{jj}$ pois $a_{jr} = 0$ se $j > r$ e $b_{rj} = 0$ se $j < r$. Segue-se imediatamente que os elementos da diagonal de $\mathbf{a}^m$ têm a forma a_{jj}^m. Daí resulta, mais geralmente, se $p(\lambda)$ é qualquer polinômio então $p(\mathbf{a})$ é uma matriz triangular superior cuja diagonal é formada pelos números $p(a_{jj})$, onde a_{jj} percorre a diagonal de $\mathbf{a}$.

Se $A \colon E \to E$ é um operador $\mathbb{C}$-linear, suas raízes características, ou seja, seus autovalores, são (contando multiplicidades) os elementos da diagonal de uma matriz triangular que representa A relativamente a uma base conveniente de E. Segue-se do que foi dito acima que, se $p(\lambda)$ é qualquer polinômio, os autovalores do operador $p(A)$, incluindo suas multiplicidades algébricas, são os números $p(\lambda_j)$, onde $\lambda_1, \ldots, \lambda_n$ são os autovalores de A.

Um caso particular interessante é o de um operador nilpotente $A \colon E \to E$. Isto significa, como se sabe, que existe um inteiro $m > 0$ tal que $A^m = 0$. Neste caso, se $n = \dim E$ afirmamos que o polinômio característico de A é $p_A(\lambda) = (-1)^n\lambda^n$.

Consideremos inicialmente o caso em que A é $\mathbb{C}$-linear. Seja $\mathcal{U} \subset E$ uma base relativamente à qual a matriz $\mathbf{a} = [a_{kj}]$ do operador A é triangular superior. Os elementos a_{jj} da diagonal de $\mathbf{a}$ são os autovalores de A, contados com suas multiplicidades. O polinômio característico de A é portanto

$$p_A(\lambda) = \prod_{j=1}^{n} (a_{jj} - \lambda).$$

Os elementos da diagonal de $\mathbf{a}^m$ são a_{jj}^m, $j = 1, \ldots, n$. Como $\mathbf{a}^m = 0$, segue-se que todos os a_{jj} são nulos, logo o polinômio característico de A é $p_A(\lambda) = (-1)^n\lambda^n$.

Do ponto de vista matricial, podemos afirmar que se $\mathbf{a}$ é uma matriz $n \times n$ (real ou complexa, tanto faz) com $\mathbf{a}^m = 0$ para algum m inteiro > 0 então seu polinômio característico é $p_{\mathbf{a}}(\lambda) = (-1)^n\lambda^n$.

Daí resulta que, dado o operador $\mathbb{R}$-linear $A \colon E \to E$ no espaço vetorial real E, com $\dim E = n$, se tivermos $A^m = 0$ para algum inteiro

$m > 0$ então o polinômio característico de A é $p_A(\lambda) = (-1)^n \lambda^n$. Com efeito, este é o polinômio característico da matriz $\mathbf{a}$ do operador A numa base qualquer de E, a qual cumpre $\mathbf{a}^m = 0$.

Em seguida, usaremos a forma triangular para provar que o determinante do operador $A_r \colon E \to E$, descomplexificado de um operador $\mathbb{C}$-linear $A \colon E \to E$, é sempre ≥ 0.

Com efeito, seja $\mathcal{U} = \{u_1, \ldots, u_n\} \subset E$ uma base na qual a matriz $\mathbf{m} = [a_{kj} + ib_{kj}]$ de A é triangular superior: $a_{kj} = b_{kj} = 0$ se $k > j$. Para todo $j = 1, \ldots, n$, temos

$$Au_j = \sum_{k=1}^{n} (a_{kj} + ib_{kj}) u_k \,.$$

Consideremos agora a base $\mathcal{U}' = \{u_1, iu_1, \ldots, u_n, iu_n\}$ do espaço vetorial real E. Para obter a matriz $\mathbf{c}$ do descomplexificado $A_r \colon E \to E$ na base $\mathcal{U}'$, observamos que

$$A_r \cdot u_j = \sum_k (a_{kj} u_j + b_{kj} \cdot iu_j) \quad \text{e} \quad A_r \cdot iu_j = \sum_k (-b_{kj} u_j + a_{kj} \cdot iu_j).$$

Daí resulta que a matriz $\mathbf{c}$ tem uma forma “triangular por blocos”.

Mostramos abaixo esta matriz no caso $n = 3$:

$$\mathbf{c} = \begin{bmatrix} a_{11} & -b_{11} & a_{12} & -b_{12} & a_{13} & -b_{13} \\ b_{11} & a_{11} & b_{12} & a_{12} & b_{13} & a_{13} \\ 0 & 0 & a_{22} & -b_{22} & a_{23} & -b_{23} \\ 0 & 0 & b_{22} & a_{22} & b_{23} & a_{23} \\ 0 & 0 & 0 & 0 & a_{33} & -b_{33} \\ 0 & 0 & 0 & 0 & b_{33} & a_{33} \end{bmatrix}.$$

Segue-se imediatamente do Teorema 19.8 que o determinante da matriz $\mathbf{c}$ é dado por:

$$\det \mathbf{c} = \prod_{j=1}^{n} \det \begin{bmatrix} a_{jj} & -b_{jj} \\ b_{jj} & a_{jj} \end{bmatrix} = \prod_{j=1}^{n} (a_{jj}^2 + b_{jj}^2). \qquad (*)$$

Sendo a matriz $\mathbf{m} = [a_{kj} + ib_{kj}]$ triangular, os números complexos $\lambda_j = a_{jj} + ib_{jj}$ são seus autovalores e $\det \mathbf{m} = \prod \lambda_j$. Por outro lado, a igualdade (*) mostra que $\det \mathbf{c} = \prod |\lambda_j|^2$.

Vemos então que $\det A_r = \det \mathbf{c} \geq 0$, valendo $\det A_r = 0$ se, e somente se, $\det A = \det \mathbf{m} = 0$. Vemos também que $\det A_r = |\det A|^2$.

Noutras palavras, dada a matriz $n \times n$ complexa $\mathbf{m} = [a_{kj} + ib_{kj}]$, formamos as matrizes reais $\mathbf{a} = [a_{kj}]$, $\mathbf{b} = [b_{kj}] \in M(n \times n)$ e

$$\mathbf{c} = \begin{bmatrix} \mathbf{a} & -\mathbf{b} \\ \mathbf{b} & \mathbf{a} \end{bmatrix} \in M(2n \times 2n).$$

Vale então $\det \mathbf{c} = |\det \mathbf{m}|^2$.

Mais uma exibição de força do Teorema 21.3 é o seu uso para demonstrar a versão geral do Teorema Espectral para operadores complexos, conforme faremos agora.

Um operador $\mathbb{C}$-linear $A\colon E \to E$ num espaço vetorial complexo, de dimensão finita, chama-se *normal* quando comuta com seu adjunto, isto é, quando cumpre a condição $AA^* = A^*A$. Analogamente, uma matriz quadrada $\mathbf{a}$ chama-se *normal* quando $\mathbf{aa}^* = \mathbf{a}^*\mathbf{a}$.

Evidentemente um operador é normal se, e somente se, sua matriz relativamente a uma (e portanto a qualquer) base ortonormal é uma matriz normal.

Teorema Espectral para operadores complexos. *Seja E um espaço vetorial complexo, munido de um produto interno hermitiano. Um operador $\mathbb{C}$-linear $A\colon E \to E$ é normal se, e somente se, existe em E uma base ortonormal formada por autovetores de A.*

Versão matricial do Teorema Espectral. *Seja a matriz $\mathbf{a} \in M(n \times n; \mathbb{C})$. A fim de que exista uma matriz unitária $\mathbf{u}$ tal que $\mathbf{d} = \mathbf{u}^*\mathbf{au}$ é diagonal é necessário e suficiente que $\mathbf{a}^*\mathbf{a} = \mathbf{aa}^*$.*

Demonstração: Continuando com a veia matricial que tem predominado nestas últimas páginas, provaremos a segunda versão, claramente equivalente à primeira. Seja $\mathbf{u}$ uma matriz unitária tal que $\mathbf{t} = \mathbf{u}^*\mathbf{au}$ seja triangular. Tomando adjuntas, vem $\mathbf{t}^* = \mathbf{u}^*\mathbf{a}^*\mathbf{u}$. Multiplicando membro a membro: $\mathbf{tt}^* = \mathbf{u}^*\mathbf{auu}^*\mathbf{a}^*\mathbf{u} = \mathbf{u}^*\mathbf{aa}^*\mathbf{u}$. Fazendo a mesma multiplicação na ordem inversa: $\mathbf{t}^*\mathbf{t} = \mathbf{u}^*\mathbf{a}^*\mathbf{au}$. Como $\mathbf{aa}^* = \mathbf{a}^*\mathbf{a}$, concluímos que $\mathbf{t}^*\mathbf{t} = \mathbf{tt}^*$. Ora, sendo triangular e normal, $\mathbf{t}$ deve ser diagonal. (compare $(\mathbf{tt}^*)_{ii}$ com $(\mathbf{t}^*\mathbf{t})_{ii}$ para $i = 1$, depois $i = 2$, etc.) Assim $\mathbf{u}^*\mathbf{au}$ é diagonal. Reciprocamente, se $\mathbf{d} = \mathbf{u}^*\mathbf{au}$ é diagonal então $\mathbf{dd}^* = \mathbf{d}^*\mathbf{d}$, o que nos dá imediatamente

$\mathbf{u}^*\mathbf{a}^*\mathbf{a}\mathbf{u} = \mathbf{u}^*\mathbf{a}\mathbf{a}^*\mathbf{u}$. Multiplicando esta igualdade à esquerda por $\mathbf{u}$ e à direita por $\mathbf{u}^*$ obtemos $\mathbf{a}^*\mathbf{a} = \mathbf{a}\mathbf{a}^*$ logo $\mathbf{a}$ é normal. □

Corolário. *Se $A: E \to E$ é um operador hermitiano, existe uma base ortonormal de E formada por autovetores de A.*

A matriz de A nesta base é diagonal e, sendo hermitiana, os elementos da diagonal são números reais. Segue-se que os autovalores de um operador hermitiano são todos reais.

Outro caso particular de operador normal é um operador unitário o qual também admite uma base ortonormal formada por auto-vetores. Os autovalores de um operador unitário são números complexos de módulo 1.

Exercícios

21.1. Seja E um espaço vetorial real. O *complexificado* de E é o conjunto E_c cujos elementos são as expressões formais $u + iv$, com $u, v \in E$ e $i = \sqrt{-1}$. Em E_c, a igualdade $u + iv = u' + iv'$ significa, por definição, que $u = u'$ e $v = v'$. A soma é definida por $(u + iv) + (u' + iv') = (u + u') + i(v + v')$ e o produto por um número complexo é, ainda por definição, $(\alpha + i\beta)(u + iv) = (\alpha u - \beta v) + i(\beta u + \alpha v)$. Para todo $u \in E$, escreve-se $u + i0 = u$ e com isso tem-se $E \subset E_c$. O complexificado de um operador linear $A: E \to E$ é $A_c: E_c \to E_c$, definido por $A_c(u + iv) = Au + iAv$. Prove:

(a) E_c é um espaço vetorial complexo e $A_c: E_c \to E_c$ é um operador $\mathbb{C}$-linear. O complexificado de $\mathbb{R}^n$ é $\mathbb{C}^n$. $E \subset E_c$ mas E não é um subespaço vetorial (complexo) de E_c.

(b) Toda $\mathbb{R}$-base $\{u_1, \ldots, u_n\} \subset E$ é uma $\mathbb{C}$-base de E_c. Em particular, $\dim_{\mathbb{C}} E_c = \dim_{\mathbb{R}} E$. A matriz de $A_c: E_c \to E_c$ relativamente à base $\{u_1, \ldots, u_n\} \subset E$ coincide com a matriz de A na mesma base. Os polinômios característicos de A_c e de A são iguais.

(c) Se $\lambda=\alpha+i\beta$, com $\beta\neq 0$, é autovalor de A_c, correspondente ao autovetor $u+iv\in E_c$, então $\{u, v\}$ é a base de um subespaço vetorial

$F \subset E$, invariante por A, e a matriz da restrição $A \colon F \to F$ é

$$\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

21.2. Seja $A \colon E \to E$ um operador linear, com $\dim E = n$. Considere o conjunto $\mathcal{M}$ de todos os polinômios $p(\lambda)$ tais que $p(A) = 0$. Prove:

(a) Se $p_1(\lambda), p_2(\lambda) \in \mathcal{M}$ e α_1, α_2 são números então $\alpha_1 p_1(\lambda) + \alpha_2 p_2(\lambda) \in \mathcal{M}$.

(b) Se $p(\lambda) \in \mathcal{M}$ e $q(\lambda)$ é qualquer polinômio então $p(\lambda)q(\lambda) \in \mathcal{M}$.

(c) Existe um único polinômio mônico $m_A(\lambda) \in \mathcal{M}$ tal que todos os outros $p(\lambda) \in \mathcal{M}$ são múltiplos de $m_A(\lambda)$. [Considere o polinômio mônico de menor grau possível em $\mathcal{M}$. Chame-o de $m_A(\lambda)$. Para todo $p(\lambda) \in \mathcal{M}$ tem-se $p(\lambda) = q(\lambda) \cdot m_A(\lambda) + r(\lambda)$, com $\text{gr} \cdot r(\lambda) < \text{gr} \cdot m_A(\lambda)$. Segue-se de (a) e (b) que $r(\lambda) \in \mathcal{M}$, logo $r(\lambda) = 0$.]

(d) O polinômio $m_A(\lambda)$ chama-se o *polinômio mínimo* do operador A. Ele é o polinômio mônico de menor grau tal que $m_A(A) = 0$. As conclusões acima valem igualmente para espaços vetoriais reais ou complexos.

(e) Se $B \colon E \to E$ é invertível então, para todo polinômio $p(\lambda)$, tem-se $p(B^{-1}AB) = B^{-1} \cdot p(A) \cdot B$. Segue-se que os operadores A e $B^{-1}AB$ têm o mesmo polinômio mínimo.

(f) Para toda matriz $\mathbf{a}$ do operador A, $m_A(\lambda)$ é o polinômio mônico de menor grau tal que $m_A(\mathbf{a}) = 0$.

21.3. Determine o polinômio mínimo dos seguintes operadores:

(a) O operador zero.

(b) O operador αI, com $\alpha \neq 0$.

(c) Uma projeção.

(d) Uma involução.

(e) O operador de derivação $D\colon \mathcal{P}_n \to \mathcal{P}_n$.

(f) O operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, $A(x,y) = (x+2y, 2x+y)$.

(g) Qualquer operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$.

Em todos estes casos, compare com o polinômio característico.

21.4. Prove que um operador é invertível se, e somente se, o termo constante do seu polinômio mínimo é $\neq 0$.

21.5. Seja $p(\lambda)$ um polinômio tal que $p(A) = 0$. Prove que todo autovalor λ_1 do operador A é raiz do polinômio $p(\lambda)$. Conclua daí que toda raiz do polinômio característico $p_A(\lambda)$ é também raiz do polinômio mínimo m_A. (A recíproca é evidente porque m_A divide p_A.)

21.6. Determine o polinômio mínimo do operador dado por $Av = \langle v, a \rangle\, b$.

21.7. Seja $A\colon E \to E$ um operador num espaço vetorial de dimensão n. Prove: se $A^k = 0$ para algum $k > n$ então $A^n = 0$.
[Sugestão: Teorema 21.3.]

21.8. Prove que um operador é nilpotente (isto é $A^k = 0$ para algum $k \in \mathbf{N}$) se, e somente se, todos os seus autovalores são iguais a zero.

21.9. Se o operador A é diagonalizável e $\lambda_1, \ldots, \lambda_k$ são seus autovalores distintos dois a dois, prove que o polinômio mínimo de A é $m_A = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_k)$.

21.10. Se o operador $A\colon E \to E$ (num espaço vetorial complexo) é diagonalizável, prove que existe um produto interno hermitiano em E que torna A normal.

21.11. Seja E um espaço vetorial (complexo) munido de um produto interno hermitiano. Prove que todo operador linear $A\colon E \to E$ se escreve, de modo único, como $A = H + iK$, onde $H, K\colon E \to E$ são operadores hermitianos e que A é normal se, e somente se, H e K comutam.

21.12. Sem fazer cálculo algum, conclua que o produto de duas matrizes $2n \times 2n$ do tipo

$$\begin{bmatrix} \mathbf{a} & -\mathbf{b} \\ \mathbf{b} & \mathbf{a} \end{bmatrix}$$

é ainda deste mesmo tipo.

21.13. Assinale V(erdadeiro) ou F(also):

( ) O determinante de um operador hermitiano é um número real.

( ) Os autovalores de um operador unitário são iguais a ± 1.

( ) Os autovalores de um operador real anti-simétrico são do tipo $\pm i\beta$, onde β é real.

( ) Se

$$\mathbf{a} = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix},$$

existe uma matriz (complexa) 2×2 invertível $\mathbf{p}$ tal que $\mathbf{p}^{-1}\mathbf{a}\mathbf{p} = \mathbf{d}$, onde $\mathbf{d}$ é diagonal (complexa).

21.14. Prove que todo operador num espaço vetorial (complexo) de dimensão n possui um subespaço invariante de dimensão $n-1$.

21.15. Sejam $\lambda_1, \ldots, \lambda_n$ os autovalores do operador $A \colon E \to E$ (repetidos de acordo com suas multiplicidades algébricas). Dado um polinômio qualquer $p(\lambda)$, prove que os autovalores do operador $p(A)$ são os números $p(\lambda_1), \ldots, p(\lambda_n)$. [Sugestão: Teorema 21.3.]

21.16. Prove que as seguintes afirmações acerca dos operadores lineares $A, B \colon E \to E$ são equivalentes:

(a) $p_A(B)$ é invertível. (Onde p_A é o polinômio característico de A.)

(b) A e B não têm autovalores em comum.

(c) $p_B(A)$ é invertível.

(d) Se m_A e m_B são os polinômios mínimos de A e B então $m_A(B)$ e $m_B(A)$ são invertíveis. [Sugestão: use o exercício anterior.]

21.17. Suponha que o polinômio mínimo $m_A(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_k)$ do operador linear $A \colon E \to E$ seja um produto de fatores distintos do primeiro grau ($\lambda_i \neq \lambda_j$ se $i \neq j$). Prove:

(a) Escrevendo $m_A(\lambda) = p_i(\lambda)(\lambda - \lambda_i)$ e $B_i = p_i(A)$, tem-se $A(B_i v) = \lambda_i B_i v$ $(i = 1, \ldots, k)$ para todo $v \in E$.

(b) Os polinômios $p_1(\lambda), \ldots, p_k(\lambda)$ são primos entre si, logo existem $q_1(\lambda), \ldots, q_k(\lambda)$ tais que

$$\sum_{i=1}^{k} q_i(\lambda) p_i(\lambda) = 1.$$

(c) Seja $C_i = q_i(A)$. Para todo $v \in E$ tem-se $v = \Sigma B_i(C_i v)$, logo os autovetores de A geram E.

(d) O operador A é diagonalizável.

21.18. Sejam A, B, X operadores lineares no espaço vetorial E. Prove:

(a) Se $AX - XB = 0$ então $A^2 X = XB^2$ e, mais geralmente, $p(A)X = Xp(B)$ para todo polinômio p.

(b) Se A e B não têm autovalores em comum então $AX = XB$ se, e somente se, $X = 0$.

(c) Se A e B não têm autovalores em comum então para todo operador linear $C \colon E \to E$ existe um único $X \in \mathcal{L}(E)$ tal que $AX - XB = C$.

21.19. Sejam $\lambda_1, \ldots, \lambda_n$ os autovalores da matriz $\mathbf{a}$, repetidos de acordo com suas multiplicidades algébricas. Prove que

$$\lambda_1^2 + \cdots + \lambda_n^2 = \sum_{i,j} a_{ij} a_{ji}.$$

21.20. Se existir algum $k \in \mathbb{N}$ tal que $A^k = I$, prove que o operador $A \colon E \to E$ é diagonalizável. [Sugestão: Exercício 21.17.]

21.21. Sejam $F_1 \subset F_2$ subespaços invariantes do operador $A \colon E \to E$. Se $\dim F_2 - \dim F_1 \geq 2$, prove que existe um subespaço F, diferente de F_1 e F_2, invariante por A, tal que $F_1 \subset F \subset F_2$.

21.22. Seja $A \colon E \to E$ um operador nilpotente no espaço vetorial E (real ou complexo) de dimensão n. Tome k, o menor número natural tal que $A^k = 0$. Prove sucessivamente:

(a) $\{0\} \subset \mathcal{N}(A) \subset \mathcal{N}(A^2) \subset \ldots \subset \mathcal{N}(A^k) = E$;

(b) Se $\mathcal{N}(A^i) = \mathcal{N}(A^{i+1})$ então $\mathcal{N}(A^{i+1}) = \mathcal{N}(A^{i+2})$;

(c) Nenhuma das inclusões em (a) se reduz a uma igualdade;

(d) $k \leq n$.

22

Equações a Diferenças Finitas

Em muitas aplicações da Matemática o tempo é discreto. Isto significa que, ao contrário da Cinemática, na qual o tempo flui continuamente, nestas situações que temos em mente (Economia, por exemplo) as grandezas são medidas em instantes isolados, formando uma seqüência. Nestes casos, as equações diferenciais não são o instrumento adequado para exprimir a evolução dos fenômenos, sendo substituídas pelas equações a diferenças finitas. Nesta seção, estudaremos os tipos mais simples dessas equações, como aplicação de alguns dos resultados obtidos nas seções anteriores.

Numa equação a diferenças finitas, a incógnita é uma seqüência $(x_0, x_1, \ldots, x_k, \ldots)$, cujos termos devem satisfazer uma relação dada. Se a relação é do tipo $x_{k+1} = f(x_k)$, onde f é uma função determinada, tem-se uma equação de primeira ordem. Se é do tipo $x_{k+2} = f(x_k, x_{k+1})$, tem-se uma equação de segunda ordem, e assim por diante.

Fixado arbitrariamente um número x_0, toda equação de primeira ordem $x_{k+1} = f(x_k)$ admite uma única solução $(x_0, x_1, \ldots, x_k, \ldots)$ com valor inicial x_0. Basta tomar sucessivamente $x_1 = f(x_0)$, $x_2 = f(x_1)$, etc. De modo análogo, fixados arbitrariamente x_0 e x_1, toda equação de segunda ordem $x_{k+2} = f(x_k, x_{k+1})$ admite uma única solução $(x_0, x_1, x_2, \ldots, x_k, \ldots)$ cujos dois valores iniciais são os núme-

ros dados. Basta tomar sucessivamente $x_2 = f(x_0, x_1)$, $x_3 = f(x_1, x_2)$, etc. E assim por diante: toda equação de ordem n possui uma única solução cujos n valores iniciais são fixados arbitrariamente.

Observação. Nesta seção, o primeiro termo de toda seqüência tem índice 0, em vez de 1.

Exemplo 22.1. A solução da equação de primeira ordem $x_{k+1} = x_k + b$ com valor inicial x_o é a seqüência $(x_o, x_o + b, x_o + 2b, \ldots)$, de termo geral $x_k = x_o + kb$ (*progressão aritmética* de razão b).

Exemplo 22.2. A equação $x_{k+1} = ax_k$ (linear, homogênea, de primeira ordem, com coeficiente constante) tem para solução, com valor inicial x_o, a seqüência $(x_o, ax_o, a^2x_o, \ldots, a^kx_o, \ldots)$ cujo termo geral é $x_k = a^k.x_o$ (*progressão geométrica* de razão a).

Exemplo 22.3. Combinando os exemplos anteriores, seja $x_{k+1} = ax_k + b$ a equação linear, não-homogênea, de primeira ordem, com coeficientes constantes. Se $(x_o, x_1, \ldots, x_k \ldots)$ é a solução desta equação com valor inicial x_o, então temos sucessivamente:

$$
\begin{aligned}
x_1 &= ax_o + b, \\
x_2 &= ax_1 + b = a^2x_o + (1 + a)b, \\
x_3 &= ax_2 + b = a^3x_o + (1 + a + a^2)b, \\
&\vdots \\
x_k &= ax_{k-1} + b = a^k.x_o + (1 + a + \cdots + a^{k-1})b
\end{aligned}
$$

Portanto a solução geral da equação $x_{k+1} = ax_k + b$ é

$$
\begin{aligned}
x_k &= a^k.x_o + \frac{1 - a^k}{1 - a} \cdot b, \qquad \text{se} \quad a \neq 1 \\
x_k &= x_o + k \cdot b, \qquad \text{se} \quad a = 1.
\end{aligned}
$$

Exemplo 22.4. A equação $x_{k+1} = ax_k + b$ pode ser olhada sob o ponto de vista de um operador linear $A \colon \mathbb{R}^\infty \to \mathbb{R}^\infty$, no espaço $\mathbb{R}^\infty$, cujos elementos são as seqüências $x = (x_o, x_1, \ldots, x_k, \ldots)$. O operador A associa a cada seqüência x a nova seqüência $y = Ax$, onde $y_k = x_{k+1} - ax_k$. A equação dada equivale ao problema de achar os elementos $x \in \mathbb{R}^\infty$ tais que $Ax = \hat{b}$, onde $\hat{b} = (b, b, \ldots)$ é uma seqüência constante de termos todos iguais a b. Como vimos no

Teorema 6.4, a solução geral da equação $Ax = \hat{b}$ é a soma de um elemento qualquer do núcleo de A (solução geral da equação homogênea $Ax=0$) com uma solução particular da equação $Ax=\hat{b}$ dada. Para obter uma dessas soluções particulares, tentamos a solução constante $\hat{c} = (c, c, \ldots)$. O número c deve cumprir $c = ac + b$, isto é, $(1-a)c = b$. Como no caso $a = 1$ já foi visto no Exemplo 22.1, supomos aqui $a \neq 1$ e obtemos $c = (1-a)^{-1}.b$. Por sua vez, a solução geral de $Ax = 0$ (equação equivalente a $x_{k+1} = ax_k$) é uma progressão geométrica $(p, ap, a^2p, \ldots)$ cujo primeiro termo p é arbitrário. Assim, a solução geral da equação $x_{k+1} = ax_k + b$, para $a \neq 1$, é dada por $x_k = a^k.p + (1-a)^{-1}.b$. Note que $x_o = p + (1-a)^{-1}.b$, donde $p = x_o - (1-a)^{-1}.b$ e, por substituição, vem:

$$x_k = a^k.x_o - a^k(1-a)^{-1}.b + (1-a)^{-1}b = a^k x_o + \frac{1-a^k}{1-a} \cdot b,$$

reobtendo o resultado do Exemplo 22.3.

22.A. Sistemas Lineares

Generalizando o Exemplo 22.2, podemos considerar, num espaço vetorial E, um operador linear $A\colon E \to E$ e procurar uma seqüência de vetores $v_k \in E$ tais que

$$v_{k+1} = A.v_k \qquad (k = 0, 1, 2, \ldots).$$

Isto se chama um sistema linear homogêneo de primeira ordem, de equações a diferenças finitas com coeficientes constantes.

Evidentemente, dado arbitrariamente um vetor inicial $v_o \in E$, existe uma única seqüência $(v_o, v_1, \ldots, v_k, \ldots)$ de vetores em E, começando com v_o e cumprindo a condição $v_{k+1} = Av_k$ para todo $k \geq 0$. Basta tomar $v_k = A^k.v_o$.

O problema prático de resolver o sistema $v_{k+1} = Av_k$ reduz-se portanto ao cálculo das potências sucessivas A^k do operador A. Em geral, isto não é uma tarefa simples. Há, entretanto, casos particulares em que ela é factível. Por exemplo, se existir uma base $\mathcal{U} \subset E$ formada por autovetores de A (o que se dá quando $\dim E = n$ e o polinômio característico de A tem n raízes reais distintas, ou então quando A é auto-adjunto), o vetor inicial v_o exprime-se como combinação linear

$$v_o = x_1 u_1 + \cdots + x_n u_n$$

de vetores $u_i \in \mathcal{U}$, com $Au_i = \lambda_i u_i$ $(i = 1, 2, \ldots, n)$. Segue-se que

$$A^k.v_o = x_1\lambda_1^k u_1 + \cdots + x_n\lambda_n^k u_n \quad (k = 0, 1, \ldots).$$

Exemplo 22.5. Seja $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ o operador linear definido por $A(x, y) = (2x + 2y, 2x - y)$. O sistema de equações a diferenças finitas $v_{k+1} = Av_k$, com $v_k = (x_k, y_k)$, escreve-se explicitamente como

$$\begin{aligned} x_{k+1} &= 2x_k + 2y_k \\ y_{k+1} &= 2x_k - y_k. \end{aligned}$$

O polinômio característico do operador A é $p(\lambda) = \lambda^2 - \lambda - 6$, cujas raízes são 3 e -2. Para obter uma base de autovetores $\{u, v\} \subset \mathbb{R}^2$, com $u = (\alpha, \beta)$ e $v = (\gamma, \delta)$, escrevemos as igualdades $Au = 3u$, $Av = -2v$ em termos de coordenadas, obtendo os sistemas

$$\begin{aligned} 2\alpha + 2\beta &= 3\alpha \\ 2\alpha - \beta &= 3\beta \\ 2\gamma + 2\delta &= -2\gamma \\ 2\gamma - \delta &= -2\delta \end{aligned}$$

Estes sistemas são indeterminados. Têm que ser indeterminados pois os números 3 e -2 foram obtidos de forma que fosse assim.

Tomemos as soluções $\alpha = 2$, $\beta = 1$, logo $u = (2, 1)$, e $\gamma = 1$, $\delta = -2$, logo $v = (1, -2)$. Então, para todo $k = 0, 1, 2, 3, \ldots$ temos

$$A^k.u = 3^k.u = (3^k.2, 3^k)$$

e

$$A^k.v = (-2)^k.v = \left((-2)^k, (-2)^{k+1}\right).$$

Para obter, digamos, a solução do sistema $v_{k+1} = Av_k$ cujo vetor inicial é $v_o = (x_o, y_o)$, com $x_o = 1$, $y_o = 1$, exprimimos o vetor $v_o = (1, 1)$ como combinação linear dos vetores básicos $u = (2, 1)$ e $v = (1, -2)$, obtendo

$$v_o = \frac{3}{5}u - \frac{1}{5}v.$$

A solução procurada é, portanto:

$$v_k = A^k.v_o = \frac{3}{5}A^k u - \frac{1}{5}A^k v = \frac{3^{k+1}}{5}u - \frac{(-2)^k}{5}v.$$

Em termos das coordenadas $v_k = (x_k, y_k)$, temos

$$x_k = \frac{2}{5}[3^{k+1} + (-2)^{k-1}] \quad \text{e} \quad y_k = \frac{1}{5}[3^{k+1} - (-2)^{k+1}].$$

No caso em que dim $E = 2$, o problema de calcular efetivamente a solução do sistema $v_{k+1} = Av_k$ pode ser resolvido em todos os casos, como mostraremos agora.

Tendo visto o caso em que o espaço E possui uma base de autovetores do operador $A\colon E \to E$, vamos agora (supondo sempre dim $E=2$) examinar os casos em que tal base não existe. Há duas possibilidades, que são as seguintes:

Primeira. O polinômio característico de A possui uma raiz real dupla λ, mas $A \neq \lambda I$. (Evidentemente, quando $A = \lambda I$ todo vetor não-nulo em E é autovetor de A, logo este caso já foi visto.)

Segunda. O polinômio característico de A possui raízes complexas $\lambda + i\mu$ e $\lambda - i\mu$ (com $i = \sqrt{-1}$ e $\mu \neq 0$).

Consideremos o primeiro destes dois casos.

Existe um vetor não-nulo $u \in E$ tal que $Au = \lambda u$. Além disso, somente os múltiplos de u podem ser autovetores de A. (Com efeito, sendo λ o único autovalor de A, se existisse um autovetor v não múltiplo de u, teríamos uma base $\{u, v\} \subset E$ com $Au = \lambda u$, $Av = \lambda v$, donde $A = \lambda I$.)

Tomemos um vetor v tal que $\{u, v\} \subset E$ seja uma base. Então $Av = \alpha u + \beta v$ com $\alpha \neq 0$ pois v não é autovetor. Se fôr $\alpha \neq 1$, substituímos v por $w = \alpha^{-1} v$ e obtemos uma nova base $\{u, w\} \subset E$ tal que $Au = \lambda u$, $Aw = u + \gamma w$. Na base $\{u, w\}$, A tem a matriz triangular

$$\begin{bmatrix} \lambda & 1 \\ 0 & \gamma \end{bmatrix},$$

cujos autovalores são λ, γ. Segue-se que $\gamma = \lambda$. Assim, temos

$$\begin{cases} Au = \lambda u \\ Aw = u + \lambda w \end{cases} \qquad (*)$$

e a matriz de A na base $\{u, w\}$ é $\mathbf{m} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$.

Das igualdades (*) resulta que para $k = 0, 1, 2, \ldots$, tem-se

$$\begin{cases} A^k u = \lambda^k u \\ A^k w = k\lambda^{k-1} u + \lambda^k w, \end{cases}$$

logo

$$\mathbf{m}^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{bmatrix}.$$

Exemplo 22.6. O operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, dado por $A(x, y) = (3x - y, x + y)$, tem o polinômio característico

$$p(\lambda) = \lambda^2 - 4\lambda + 4,$$

o qual admite a raiz dupla λ=2, sendo u=$(1,1)$ um autovetor de A. Portanto $\{u, e_2\} \subset \mathbb{R}^2$ é uma base, com $Au = 2u$ e $Ae_2 = (-1, 1) = -1 \cdot u + 2 \cdot e_2$. Tomando $w = -e_2 = (0, -1)$, obtemos a base $\{u, w\} \subset \mathbb{R}^2$, com $Au = 2u$ e $Aw = u + 2w$, logo a matriz de A na base $\{u, w\}$ é

$$\mathbf{m} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}.$$

Pelo que foi dito acima, para obter a solução $v_k = (x_k, y_k) = A^k.v_o$ do sistema

$$\begin{aligned} x_{k+1} &= 3x_k - y_k, \\ y_{k+1} &= x_k + y_k, \end{aligned}$$

com vetor inicial $v_o = (3, 5)$, primeiro exprimimos v_o como combinação linear de u e w, obtendo $v_o = 3u - 2w$. Daí resulta que, para todo $k = 0, 1, 2, \ldots$, tem-se

$$v_k = 3.A^k u - 2.A^k w.$$

Como vimos acima, $A^k u = 2^k u$ e $A^k w = k2^{k-1} u + 2^k w$. Logo

$$v_k = 2^k[(3 - k)u - 2w] = 2^k(3 - k, 5 - k).$$

Em termos das coordenadas, isto significa que

$$x_k = 2^k(3 - k) \quad \text{e} \quad y_k = 2^k(5 - k).$$

Para completar o estudo de sistemas lineares de equações a diferenças finitas no caso 2×2, trataremos agora o caso de um operador linear $A\colon E \to E$, (com dim $E = 2$) cujas raízes características são números complexos $\lambda + i\mu, \lambda - i\mu$, com $\mu \neq 0$.

Será conveniente considerar o *complexificado* de E, que é o espaço vetorial complexo E_c, de dimensão (complexa) 2, cujos elementos têm a forma $u + iv$, onde $u, v \in E$. As operações em E_c são dadas por

$$(u + iv) + (u' + iv') = (u + u') + i(v + v')$$

e

$$(\alpha + i\beta)(u + iv) = (\alpha u - \beta v) + i(\alpha v + \beta \mu).$$

Tem-se $E \subset E_c$ de modo natural, pois $u = u + i.0$. O operador $A\colon E \to E$ estende-se a um operador $A_c\colon E_c \to E_c$, chamado o *complexificado* de A, pondo-se, por definição, $A_c(u + iv) = Au + iAv$. Toda base de E é também uma base de E_c, relativamente à qual a matriz de A_c é a mesma matriz de A. Em particular, os polinômios característicos de A e A_c coincidem, logo $\lambda + i\mu$ e $\lambda - i\mu$ são autovalores distintos do operador $\mathbb{C}$-linear $A_c\colon E_c \to E_c$. Para nossos efeitos, basta considerar o autovalor $\lambda + i\mu$.

Seja $u + iv \in E_c$ um autovetor de A_c correspondente ao autovalor $\lambda + i\mu$. Então $u, v \in E$ são vetores não simultaneamente nulos, com $A_c(u + iv) = (\lambda + i\mu)(u + iv)$, ou seja:

$$Au + iAv = (\lambda u - \mu v) + i(\mu u + \lambda v),$$

logo

$$Au = \lambda u - \mu v \quad \text{e} \quad Av = \mu u + \lambda v. \tag{*}$$

Afirmamos que os vetores $u, v \in E$ são linearmente independentes. Em primeiro lugar, u e v são ambos $\neq 0$ pois se um deles fosse nulo o outro seria $\neq 0$ e, pelas equações (*), este seria um autovetor do operador $A\colon E \to E$. Em seguida, se u e v fossem L.D. teríamos $v = \alpha u$, logo $Au = \lambda u - \mu v = (\lambda - \alpha\mu)u$ e u seria um autovetor de A.

Portanto, $\{u, v\} \subset E$ é uma base, relativamente à qual a matriz do operador $A\colon E \to E$ tem a forma

$$\mathbf{a} = \begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix}.$$

O cálculo das potências do operador A ou, equivalentemente, da matriz **a** se baseia na observação de que as matrizes da forma

$$\begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix}$$

se comportam, em relação à adição e à multiplicação, do mesmo modo que os números complexos $\lambda + i\mu$.

Mais precisamente, se associarmos a cada número complexo $z = x + iy$ a matriz

$$\varphi(z) = \begin{bmatrix} x & y \\ -y & x \end{bmatrix}$$

teremos uma correspondência injetiva φ tal que $\varphi(z+w) = \varphi(z) + \varphi(w)$ e $\varphi(z.w) = \varphi(z).\varphi(w)$, onde $z+w$ e $z.w$ são as operações de adição e multiplicação de números complexos e $\varphi(z) + \varphi(w)$ e $\varphi(z).\varphi(w)$ são as operações correspondentes com matrizes.

Assim, se quisermos calcular a k-ésima potência da matriz

$$\begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix}$$

basta calcular $(\lambda + i\mu)^k = x + iy$ e teremos

$$\begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix}^k = \begin{bmatrix} x & y \\ -y & x \end{bmatrix}.$$

Ora, as potências de um número complexo se calculam facilmente com a fórmula de De Moivre: basta escrevê-lo sob a forma trigonométrica $\lambda + i\mu = \rho(\cos\theta + i\,\text{sen}\,\theta)$ e então $(\lambda+i\mu)^k = \rho^k(\cos k\theta + i\,\text{sen}\,k\theta)$. Aí, $\rho = \sqrt{\lambda^2 + \mu^2}$, $\cos\theta = \lambda/\rho$ e $\text{sen}\,\theta = \mu/\rho$. Portanto:

$$\begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix}^k = \rho^k \begin{bmatrix} \cos k\theta & \text{sen}\,k\theta \\ -\text{sen}\,k\theta & \cos k\theta \end{bmatrix}.$$

Se temos, por exemplo, um operador $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, dado por $A(x,y) = (ax+by, cx+dy)$, cujas raízes características são os números complexos $\lambda \pm i\mu$, segue-se da discussão acima que existe uma base $\{u, v\} \subset \mathbb{R}^2$ em relação à qual a matriz de A^k tem a forma

$$\rho^k \begin{bmatrix} \cos k\theta & \text{sen}\,k\theta \\ -\text{sen}\,k\theta & \cos k\theta \end{bmatrix},$$

onde

$$\rho = \sqrt{\lambda^2 + \mu^2}, \quad \cos\theta = \lambda/\rho, \quad \operatorname{sen}\theta = \mu/\rho.$$

Depois de obtidos λ, μ, a base $\{u, v\}$ se determina resolvendo o sistema $Au = \lambda u - \mu v$, $Av = \mu u + \lambda v$, de 4 equações com 4 incógnitas que são as coordenadas dos vetores $u = (x, y)$ e $v = (s, t)$. Em termos dessas coordenadas, o sistema se escreve como

$$\begin{aligned} ax + by &= \lambda x - \mu s \\ cx + dy &= \lambda y - \mu t \\ as + bt &= \mu x + \lambda s \\ cs + dt &= \mu y + \lambda t \end{aligned}$$

ou

$$\begin{array}{lllll} (a-\lambda)x & +by & +\mu s & & = 0 \\ cx & +(d-\lambda)y & & +\mu t & = 0 \\ -\mu x & & +(a-\lambda)s & +bt & = 0 \\ & -\mu y & +cs & +(d-\lambda)t & = 0. \end{array}$$

Exemplo 22.7. Consideremos o sistema $x_{k+1} = 3x_k - y_k$, $y_{k+1} = 2x_k + y_k$, com vetor inicial $v_o = (1, 1)$. Ele nos fornece o operador $A \colon \mathbb{R}^2 \to \mathbb{R}^2$, $A(x, y) = (3x - y, 2x + y)$, cuja matriz na base canônica é

$$\begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix},$$

logo o polinômio característico é $p(\lambda) = \lambda^2 - 4\lambda + 5$, cujas raízes são $2 \pm i$. Existe uma base $\{u, v\} \subset \mathbb{R}^2$, com $u = (x, y)$, $v = (s, t)$, tal que $Au = 2u - v$, $Av = u + 2v$. Para obter esta base, devemos achar uma solução não-nula do sistema

$$\begin{aligned} 3x - y &= 2x - s \\ 2x + y &= 2y - t \\ 3s - t &= x + 2s \\ 2s + t &= y + 2t \end{aligned}$$

ou seja

$$\begin{array}{rrrl} x - y & +s & & = 0 \\ 2x - y & & +t & = 0 \\ -x & +s & -t & = 0 \\ -y & +2s & -t & = 0. \end{array}$$

Por escalonamento, encontramos $x = s - t$, $y = 2s - t$, onde s, t são arbitrários. Tomamos $s = 1$, $t = 0$ e obtemos $x = 1$, $y = 2$. Logo $u = (1, 2)$ e $v = (1, 0)$ formam uma base de $\mathbb{R}^2$, relativamente à qual o operador A tem a matriz

$$\mathbf{a} = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}.$$

O número complexo $2 + i$ tem módulo $\rho = \sqrt{5}$. Um ângulo θ tal que $\cos\theta = 2/\sqrt{5}$ é $\theta = 26°33'54''$ ou seja, $\theta = 0,463$ rad. Então, para todo $k = 0, 1, 2 \ldots$, temos

$$\mathbf{a}^k = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}^k = (\sqrt{5})^k \begin{bmatrix} \cos k\,\theta & \operatorname{sen} k\,\theta \\ -\operatorname{sen} k\,\theta & \cos k\,\theta \end{bmatrix}.$$

A solução do sistema $v_{k+1} = Av_k$ com vetor inicial $v_o = (1, 1)$ é dada por $v_k = A^k.v_o$. Para obtê-la explicitamente, devemos começar exprimindo v_o como combinação linear dos vetores básicos $u = (1, 2)$ e $v = (1, 0)$. Temos $v_o = \frac{1}{2}u + \frac{1}{2}v$. Portanto

$$v_k = \frac{1}{2}A^k u + \frac{1}{2}A^k v.$$

Ora, $\mathbf{a}^k$ é a matriz de A^k na base $\{u, v\}$. Logo

$$A^k u = 5^{k/2}(\cos k\,\theta \cdot u - \operatorname{sen} k\,\theta \cdot v)$$

e

$$A^k v = 5^{k/2}(\operatorname{sen} k\,\theta \cdot u + \cos k\,\theta \cdot v).$$

Noutras palavras:

$$A^k u = 5^{k/2}(\cos k\,\theta - \operatorname{sen} k\,\theta, 2\cos k\,\theta)$$

e

$$A^k v = 5^{k/2}(\operatorname{sen} k\,\theta + \cos k\,\theta, 2\operatorname{sen} k\,\theta).$$

Portanto

$$v_k = \frac{1}{2}A^k u + \frac{1}{2}A^k v = 5^{k/2}(\cos k\,\theta, \cos k\,\theta + \operatorname{sen} k\,\theta).$$

Concluímos então que

$$x_k = 5^{k/2}\cos k\,\theta;$$

$$y_k = 5^{k/2}(\cos k\,\theta + \operatorname{sen} k\,\theta),$$

com $\theta = 26°33'54'' = 0{,}463$ rad, é a solução do sistema $x_{k+1} = 3x_k - y_k$, $y_{k+1} = 2x_k + y_k$ que cumpre $x_o = 1$, $y_o = 1$.

22.B. Uma Aplicação do Teorema de Cayley-Hamilton

Uma forma alternativa de calcular as potências sucessivas A^k do operador linear $A\colon E \to E$ com $\dim E = n$, consiste em observar que, em conseqüência do Teorema de Cayley-Hamilton, basta que se considerem os expoentes $k \le n-1$. Com efeito, se

$$p_A(\lambda) = (-1)^n \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_o$$

é o polinômio característico do operador linear A, segue-se de $p_A(A) = 0$ que

$$A^n = \pm(a_{n-1}A^{n-1} + \cdots + a_1 A + a_o I).$$

Usaremos este fato para calcular as potências A^k de um operador linear $A\colon E \to E$, onde $\dim E = 2$. Neste caso, o grau do polinômio característico $p_A(\lambda)$ sendo igual a 2, segue-se que o resto da divisão de λ^k por $p_A(\lambda)$, para todo $k \ge 2$, tem a forma $\alpha\lambda + \beta$. (Por simplicidade, escrevemos α, β em vez de α_k, β_k.) Podemos portanto escrever

$$\lambda^k = p_A(\lambda)\cdot q(\lambda) + \alpha\lambda + \beta,$$

donde

$$A^k = p_A(A)\cdot q(A) + \alpha A + \beta I.$$

Como $p_A(A) = 0$, segue-se que $A^k = \alpha A + \beta I$.

Para encontrar α e β, suponhamos inicialmente que as raízes características λ_1 e λ_2 sejam distintas. Por definição, temos $p_A(\lambda_1) = p_A(\lambda_2) = 0$ logo da identidade $\lambda^k = p_A(\lambda)\cdot q(\lambda) + \alpha\lambda + \beta$ resultam as igualdades:

$$\alpha\lambda_1 + \beta = \lambda_1^k$$
$$\alpha\lambda_2 + \beta = \lambda_2^k.$$

Como estamos supondo $\lambda_1 \neq \lambda_2$, temos acima um sistema determinado, que nos permite obter valores únicos para α e β. Observe que este argumento vale, inclusive, quando as raízes características λ_1 e λ_2 são números complexos. Neste caso, usa-se a forma trigonométrica para calcular as potências λ_1^k e λ_2^k e vê-se que as soluções α, β do sistema acima são números reais.

Consideremos agora o caso em que o polinômio característico $p_A(\lambda)$ do operador linear A possui uma raiz (real) dupla λ_1. Então

sabemos que λ_1, além de raiz do polinômio $p_A(\lambda)$, é também raiz de sua derivada $p'_A(\lambda)$. Portanto a identidade

$$\lambda^k = p_A(\lambda) \cdot q(\lambda) + \alpha\lambda + \beta$$

fornece a equação $\alpha\lambda_1 + \beta = \lambda_1^k$ enquanto que sua derivada

$$k\lambda^{k-1} = p'_A(\lambda) \cdot q(\lambda) + p_A(\lambda) \cdot q'(\lambda) + \alpha$$

fornece, em virtude das relações $p'_A(\lambda_1) = 0$ e $p_A(\lambda_1) = 0$, a igualdade $\alpha = k \cdot \lambda_1^{k-1}$, logo $\beta = \lambda_1^k(1-k)$.

22.C. Equações Lineares de Segunda Ordem

Estudaremos apenas as que têm coeficientes constantes. Primeiro as homogêneas:

$$x_{k+2} + ax_{k+1} + bx_k = 0. \tag{*}$$

Podemos reduzir o estudo da equação acima ao sistema linear de primeira ordem:

$$\begin{aligned} x_{k+1} &= y_k \\ y_{k+1} &= -bx_k - ay_k, \end{aligned} \tag{**}$$

onde foi introduzida a incógnita auxiliar $y_k = x_{k+1}$. (Em particular, $y_o = x_1$.) Usando os métodos do item 22.A, obtemos as seqüências (x_k) e (y_k), com valores iniciais x_o, y_o dados. Isto significa que a equação (*), quando são fixados os valores iniciais x_o, x_1, tem por solução a seqüência (x_k) que responde ao sistema (**). Com efeito, temos:

$$x_{k+2} = y_{k+1} = -bx_k - ay_k = -bx_k - ax_{k+1}.$$

Portanto $x_{k+2} + ax_{k+1} + bx_k = 0$.

Observe que o polinômio característico do sistema (**) é $p(\lambda) = \lambda^2 + a\lambda + b$. Isto sugere que, a fim de estudar a equação (*), não é necessário ter feito anteriormente o estudo dos sistemas lineares.

A seguir, mostraremos como resolver a equação (*), independentemente de sistemas lineares.

A primeira observação a fazer é que o subconjunto $S \subset \mathbb{R}^\infty$, formado pelas seqüências $x = (x_o, x_1, \ldots, x_k, \ldots)$ que são soluções da equação $x_{k+2} + ax_{k+1} + bx_k = 0$, é um subespaço vetorial de dimensão 2.

O fato de que S é um subespaço vetorial é de verificação imediata. (S é o núcleo do operador linear $A\colon \mathbb{R}^\infty \to \mathbb{R}^\infty$, definido por $Ax = y$, onde $y_k = x_{k+2} + ax_{k+1} + bx_k$.) Além disso, o comentário feito no início desta seção sobre a existência e unicidade da solução de uma equação de segunda ordem $x_{k+2} = f(x_k, x_{k+1})$, com valores iniciais x_o, x_1 pré-fixados, significa precisamente que a correspondência $S \to \mathbb{R}^2$, que associa a cada solução $x = (x_k)$ da equação (*) seus dois primeiros termos (x_o, x_1), nesta ordem, é um isomorfismo entre S e $\mathbb{R}^2$. Portanto o espaço vetorial S tem dimensão 2.

Este argumento mostra ainda que se $x = (x_k)$ e $x' = (x'_k)$ são duas soluções da equação $x_{k+2} + ax_{k+1} + bx_k = 0$ tais que os vetores (x_o, x_1) e (x'_o, x'_1) são linearmente independentes então toda solução desta equação se exprime, de modo único, como combinação linear $\alpha x + \beta x'$.

A segunda observação é que se r é uma raiz do polinômio característico $\lambda^2 + a\lambda + b = 0$ então a seqüência $r^* = (1, r, r^2, \ldots, r^k, \ldots)$ é uma solução da equação $x_{k+2} + ax_{k+1} + bx_k = 0$.

Com efeito, de $r^2 + ar + b = 0$ segue-se que

$$r^{k+2} + ar^{k+1} + br^k = r^k(r^2 + ar + b) = r^k \times 0 = 0.$$

Resulta dessas duas observações que, para determinar todas as soluções da equação $x_{k+2} + ax_{k+1} + bx_k = 0$, devemos usar as raízes do seu polinômio característico a fim de obter duas soluções linearmente independentes. Todas as demais soluções serão combinações lineares destas.

Há 3 casos a considerar.

Primeiro caso. O polinômio característico $\lambda^2 + a\lambda + b$ tem duas raízes reais distintas r, s.

Então as seqüências

$$r^* = (1, r, r^2, \ldots, r^k, \ldots) \quad \text{e} \quad s^* = (1, s, s^2, \ldots, s^k, \ldots)$$

são soluções e, como $r \neq s$, os vetores inicias $(1, r)$ e $(1, s)$ são L.I. em $\mathbb{R}^2$, logo r^* e s^* são linearmente independentes em $\mathbb{R}^\infty$. A solução geral da equação $x_{k+2} + ax_{k+1} + bx_k = 0$ é, portanto,

$$x_k = \alpha r^k + \beta s^k,$$

onde as constantes α e β podem ser determinadas de modo que x_o e x_1 tenham valores pré-fixados.

Exemplo 22.8. A equação $x_{k+2} - 3x_{k+1} + 2x_k = 0$ tem o polinômio característico $\lambda^2 - 3\lambda + 2$, cujas raízes são $r = 1$, $s = 2$. A solução geral desta equação tem a forma $x_k = \alpha + 2^k.\beta$. Se quisermos, por exemplo, a solução com $x_o = 1$ e $x_1 = 0$, temos que achar α, β tais que $\alpha + \beta = 1$ e $\alpha + 2\beta = 0$, o que nos dá $\alpha = 2$, $\beta = -1$, logo a solução procurada tem a forma $x_k = 2 - 2^k$.

Segundo caso. O polinômio característico $\lambda^2 + a\lambda + b$ tem uma raiz real dupla $r \neq 0$.

Tem-se $r = -a/2$, logo $2r + a = 0$. Já sabemos que uma solução da equação $x_{k+2} + ax_{k+1} + bx_k = 0$ é $r^* = (1, r, \ldots, r^k, \ldots)$. Afirmamos que $r^{**} = (0, r, 2r^2, \ldots, kr^k, \ldots)$ é outra solução. Com efeito, se $x_k = kr^k$ então

$$\begin{aligned} x_{k+2} + ax_{k+1} + bx_k &= (k+2)r^{k+2} + a(k+1)r^{k+1} + bkr^k \\ &= r^k[k(r^2 + ar + b) + r(2r + a)] = 0. \end{aligned}$$

Além disso, como os vetores $(1, r)$ e $(0, r)$ são L.I. em $\mathbb{R}^2$, segue-se que r^* e r^{**} são soluções linearmente independentes, logo a solução geral da equação dada tem a forma

$$x_k = \alpha r^k + \beta k r^k = r^k(\alpha + \beta k),$$

onde as constantes α e β podem ser determinadas de maneira a fazer com que x_o e x_1 assumam os valores iniciais pré-estabelecidos.

Exemplo 22.9. Seja a equação $x_{k+2} - 6x_{k+1} + 9x_k = 0$. Seu polinômio característico tem a raiz dupla $r = 3$. A solução geral desta equação é $x_k = 3^k(\alpha + \beta k)$. Se impusermos os valores iniciais $x_o = -1$, $x_1 = 1$ obteremos $\alpha = -1$, $\beta = 4/3$, logo a solução que tem esses valores iniciais é $x_k = 3^k(-1 + 4k/3)$.

Exemplo 22.10. Uma progressão aritmética pode também ser considerada como solução de uma equação a diferenças finitas de segunda ordem, a saber, a equação $x_{k+2} - x_{k+1} = x_{k+1} - x_k$. Escrevendo-a sob a forma $x_{k+2} - 2x_{k+1} + x_k = 0$, vemos que seu polinômio característico possui a raiz dupla $r = 1$, logo sua solução geral é $x_k = \alpha + \beta k$, ou seja, é a progressão aritmética de primeiro termo α e razão β.

Terceiro caso. As raízes do polinômio característico $\lambda^2 + a\lambda + b$ são os números complexos $\alpha \pm i\beta$ com $\beta \neq 0$, $i = \sqrt{-1}$.

Escrevemos o número complexo $r = \alpha + i\beta$ sob a forma trigonométrica $r = \rho(\cos\theta + i\,\text{sen}\,\theta)$. Exatamente como no caso real, verificamos que a seqüência de números complexos $r^k = \rho^k(\cos k\theta + i\,\text{sen}\,k\theta)$, $k = 0, 1, 2, \ldots$, é solução da equação dada. Mas é óbvio que se a, b são números reais e a seqüência complexa $(z_o, z_1, \ldots, z_k, \ldots)$, com $z_k = x_k + iy_k$, é solução da equação $z_{k+2} + az_{k+1} + bz_k = 0$ então sua parte real x_k e sua parte imaginária y_k cumprem $x_{k+2} + ax_{k+1} + bx_k = 0$ e $y_{k+2} + ay_{k+1} + by_k = 0$ respectivamente.

Logo as seqüências $x_k = \rho^k \cos k\theta$ e $y_k = \rho^k\,\text{sen}\,k\theta$ são soluções da equação dada. Além disso, como $(x_o, x_1) = (1, \rho\cos\theta)$, $(y_o, y_1) = (0, \rho\,\text{sen}\,\theta)$ e $\text{sen}\,\theta \neq 0$ (pois o número r não é real), vemos que (x_o, x_1) e (y_o, y_1) formam uma base de $\mathbb{R}^2$, logo as soluções $x_k = \rho^k \cos k\theta$ e $y_k = \rho^k\,\text{sen}\,k\theta$ são linearmente independentes.

Segue-se que a solução geral da equação $x_{k+2} + ax_{k+1} + bx_k = 0$, quando $a^2 < 4b$, é dada por

$$x_k = \rho^k[\alpha\cos k\theta + \beta\,\text{sen}\,k\theta],$$

onde as constantes α, β podem ser determinadas de modo que x_o e x_1 assumam valores arbitrariamente pré-fixados.

Exemplo 22.11. O polinômio característico da equação $x_{k+2} - 2x_{k+1} + 4x_k = 0$ tem as raízes complexas $1 \pm i\sqrt{3}$. Escrevendo $1 + i\sqrt{3}$ sob forma trigonométrica, temos, $1 + i\sqrt{3} = 2(\cos 60^\circ + i\,\text{sen}\,60^\circ) = 2\left(\cos\frac{\pi}{3} + i\,\text{sen}\,\frac{\pi}{3}\right)$. A solução geral da equação dada é

$$x_k = 2^k\left[\alpha\cos\frac{k\pi}{3} + \beta\,\text{sen}\,\frac{k\pi}{3}\right].$$

Se quisermos, por exemplo, obter a solução tal que $x_o = 5$, $x_1 = 7$, os números α, β devem satisfazer as condições $\alpha = 5$, $\beta = 2\sqrt{3}/3$.

Como se vê, a discussão direta das equações de segunda ordem (lineares homogêneas, com coeficientes constantes) é bem mais simples do que o estudo dos sistemas dos quais elas são casos particulares.

22.D. Equações com Segundo Membro Constante

As equações de que vamos tratar são as do tipo

$$x_{k+2} + ax_{k+1} + bx_k = c.$$

Como se sabe, a solução geral desta equação é a soma de uma solução particular, que obtenhamos por qualquer processo, com a solução geral da equação homogênea $x_{k+2} + ax_{k+1} + bx_k = 0$. Como já aprendemos a determinar esta última, basta-nos explicar como se pode conseguir uma solução particular.

Começamos tentando uma solução constante $x_k = d$. Devemos ter $d + ad + bd = c$, ou seja $(1+a+b)d = c$. Portanto, se $1+a+b \neq 0$, a única solução constante possível para a equação dada é $x_k = d = (1+a+b)^{-1}.c$.

Se $1 + a + b = 0$, a equação dada (com $c \neq 0$) não pode ter solução constante. Tentamos uma solução do tipo $x_k = k.d$. Substituindo, na equação dada, x_k por kd vem

$$(k+2)d + a(k+1)d + bkd = c$$

ou

$$(1+a+b)kd + (a+2)d = c.$$

Como $1+a+b = 0$, obtemos $(a+2)d = c$. Portanto, quando $1+a+b = 0$ e $a \neq -2$, a seqüência $x_k = kc.(a+2)^{-1}$ é uma solução particular.

Finalmente, se $1+a+b = 0$ e $a = -2$ então $b = 1$ e a equação dada se torna $x_{k+2} - 2x_{k+1} + x_k = c$, a qual não possui solução constante nem do tipo $x_k = k.d$. Tentemos uma solução da forma $x_k = k^2.d$. Substituindo x_k por este valor na equação, obtemos

$$(k^2 + 4k + 4)d - 2(k^2 + 2k + 1)d + k^2 d = c,$$

ou seja, $2d = c$, donde $d = c/2$. Uma verificação imediata mostra que, realmente, $x_k = k^2 c/2$ é uma solução particular da equação $x_{k+2} - 2x_{k+1} + x_k = c$.

Observação. No início desta seção, vimos que a equação de segunda ordem $x_{k+2} + ax_{k+1} + bx_k = 0$ pode ser resolvida considerando-se o sistema

$$\begin{aligned} x_{k+1} &= y_k \\ y_{k+1} &= -bx_k - ay_k. \end{aligned}$$

Mostraremos agora que, reciprocamente, se quisermos resolver o sistema

$$\begin{aligned} x_{k+1} &= ax_k + by_k \\ y_{k+1} &= cx_k + dy_k \end{aligned} \tag{*}$$

com vetor inicial $v_0 = (x_0, y_0)$, podemos reduzi-lo a uma equação de segunda ordem, resolver essa equação pelo método que acabamos de expor e, a partir daí, obter a solução do sistema.

Evidentemente, um dos números a, b, c, d é diferente de zero. Para fixar idéias, suporemos que $b \neq 0$. Se (x_k) e (y_k) são as soluções do sistema (∗) então

$$\begin{aligned} x_{k+2} &= ax_{k+1} + by_{k+1} \\ &= ax_{k+1} + bcx_k + bdy_k \\ &= ax_{k+1} + bcx_k + dx_{k+1} - adx_k, \end{aligned}$$

logo,

$$x_{k+2} - (a+d)x_{k+1} + (ad - bc)x_k = 0. \qquad (**)$$

Para resolver o sistema (∗) com vetor inicial $v_0 = (x_0, y_0)$, tomamos a solução (x_k) da equação (∗∗) com os valores iniciais x_0 (dado) e $x_1 = ax_0 + by_0$ (com y_0 também dado). Em seguida, definimos a seqüência (y_k) pondo $y_k = (x_{k+1} - ax_k)/b$. Isto dá imediatamente $x_{k+1} = ax_k + by_k$. Além disso, o valor y_0 obtido nesta fórmula coincide com o valor inicial y_0 anteriormente estipulado. Tem-se ainda

$$\begin{aligned} y_{k+1} &= \frac{1}{b}[x_{k+2} - ax_{k+1}] \\ &= \frac{1}{b}[(a+d)x_{k+1} + (bc - ad)x_k - ax_{k+1}] \\ &= \frac{1}{b}[(a+d)(ax_k + by_k) + (bc - ad)x_k - ax_{k+1}]. \end{aligned}$$

Simplificando, vem $y_{k+1} = cx_k + dy_k$, logo as seqüências (x_k) e (y_k) formam a solução procurada do sistema (∗).

Exercícios

22.1. Para cada uma das equações abaixo, determine a solução que tem o valor inicial indicado

(a) $x_{k+1} = x_k - 7$, $x_o = 0$.

(b) $x_{k+1} = 6x_k$, $x_o = 1$.

(c) $x_{k+1} = 2x_k + 5$, $x_o = 3$.

(d) $x_{k+1} = x_k + k$, $x_o = 2$.

(e) $x_{k+1} = \frac{k+1}{k+2} \cdot x_k$, $x_o = x_o$.

22.2. Prove que o conjunto das soluções de uma equação do tipo $x_{k+1} = a_k x_k + b_k$ (linear, de primeira ordem, homogênea ou não, com coeficientes variáveis) é uma variedade afim de dimensão 1 em $\mathbb{R}^\infty$. Em que condições é um subespaço vetorial?

22.3. Uma solução do sistema

$$\begin{aligned} x_{k+1} &= a_k x_k + b_k y_k + p_k \\ y_{k+1} &= c_k x_k + d_k y_k + q_k \end{aligned}$$

pode ser considerada como um par (s, t) de seqüências

$$\begin{aligned} s &= (x_o, \ldots, x_k, \ldots), \\ t &= (y_o, \ldots, y_k, \ldots), \end{aligned}$$

portanto um elemento de $\mathbb{R}^\infty \times \mathbb{R}^\infty$. Prove que o conjunto S dessas soluções é uma variedade afim de dimensão 2 no espaço vetorial $\mathbb{R}^\infty \times \mathbb{R}^\infty$. Em que condições S é um subespaço vetorial?

22.4. Para cada um dos sistemas abaixo, determine a solução $(v_o, \ldots, v_k, \ldots)$, $v_k = (x_k, y_k)$, que tem o valor inicial $v_o = (x_o, y_o)$ indicado.

$$\begin{array}{lll} x_{k+1} = 2x_k + 9y_k & x_{k+1} = 3x_k + 16y_k & x_{k+1} = x_k + 3y_k \\ y_{k+1} = x_k + 2y_k & y_{k+1} = -4x_k - 13y_k & y_{k+1} = -2x_k + y_k \\ (x_o, y_o) = (3, -2) & (x_o, y_o) = (3, 2) & (x_o, y_o) = (-2, 3) \end{array}$$

22.5. Seja $v_k = (x_k, y_k)$ uma solução do sistema

$$\begin{aligned} x_{k+1} &= 2x_k - y_k \\ y_{k+1} &= 4x_k - 2y_k \,. \end{aligned}$$

Prove que, seja qual for o vetor inicial $v_o = (x_o, y_o)$, tem-se $x_k = y_k = 0$ para todo $k \geq 2$.

22.6. Sejam λ e μ as raízes características (reais ou complexas) do operador $A: \mathbb{R}^2 \to \mathbb{R}^2$. Suponha que $|\lambda| < 1$ e $|\mu| < 1$. Seja qual for o valor inicial $v_o = (x_o, y_o)$, prove que a solução $v_k = (x_k, y_k)$ do sistema $v_{k+1} = Av_k$ cumpre $\lim x_k = 0$, $\lim y_k = 0$.

22.7. Sejam $r = (x_o, \ldots, x_k, \ldots)$ e $s = (y_o, \ldots, y_k, \ldots)$ soluções da equação $z_{k+2} + az_{k+1} + bz_k = 0$. Assinale V(erdadeiro) ou F(also):

( ) Se r e s são L.I. então, para quaisquer índices $k \neq \ell$, os vetores $u = (x_k, x_\ell)$ e $v = (y_k, y_\ell)$ são L.I.

( ) Se existirem $k \neq \ell$ tais que os vetores $u = (x_k, x_\ell)$ e $v = (y_k, y_\ell)$ são L.I. então as soluções r e s são L.I.

( ) Se, para todo $k \geq 0$, os vetores $u = (x_k, x_{k+1})$ e $v = (y_k, y_{k+1})$ forem L.D. então r e s são L.D.

( ) Se r e s são L.D. então $u = (x_k, x_\ell)$ e $v = (y_k, y_\ell)$ são L.D., sejam quais forem k e ℓ.

22.8. Sejam $r=(x_o, \ldots, x_k, \ldots)$, $s=(y_o, \ldots, y_k, \ldots)$, $t=(z_o, \ldots, z_k, \ldots)$ soluções da equação $w_{k+3} + aw_{k+2} + bw_{k+1} + c_k w_k = 0$. Prove:

(a) Se existirem $k < \ell < m$ tais que os vetores $v = (x_k, x_\ell, x_m)$, $v' = (y_k, y_\ell, y_m)$ e $v'' = (z_k, z_\ell, z_m)$ são L.I. então as seqüências r, s e t são L.I. .

(b) Se r, s e t são L.I. então existe $k \geq 0$ tal que os vetores

$$v=(x_k, x_{k+1}, x_{k+2}), \quad v'=(y_k, y_{k+1}, y_{k+2}) \quad \text{e} \quad v''=(z_k, z_{k+1}, z_{k+2})$$

são L.I. .

22.9. Prove que as seqüências $r = (1, 2, 3, 4, 0, 0, \ldots)$, $s = (1, 2, 3, 1, 0, 0, \ldots)$ e $t = (1, 0, 0, 3, 0, 0, \ldots)$ não podem ser soluções da mesma equação $x_{k+3} + ax_{k+2} + bx_{k+1} + cx_k = 0$. [Sugestão: item (b) do exercício anterior.]

22.10. Seja $E: \mathbb{R}^\infty \to \mathbb{R}^\infty$ o operador linear definido por

$$E(x_o, x_1, \ldots, x_k, \ldots) = (x_1, x_2, \ldots, x_{k+1}, \ldots).$$

Prove que um subespaço vetorial $S \subset \mathbb{R}^\infty$ é o conjunto das soluções de uma equação linear homogênea de ordem n com coeficientes constantes,

$$x_{k+n} + a_1 x_{k+n-1} + \cdots + a_n x_k = 0,$$

se, e somente se, cumpre as condições seguintes:

(1) S é invariante por E;

(2) S tem dimensão finita, igual a n.

22.11. O método apresentado em 22.B para calcular as potências A^k mediante o uso do Teorema de Cayley-Hamilton dá fórmulas explícitas para essas potências como combinações lineares de $I, A, \ldots, A^{n-1}$, mas faz uso das raízes características do operador A, que são facilmente calculáveis em dimensão 2 porém difíceis de obter em dimensões superiores. Caso se deseje, não a fórmula geral para A^k, mas apenas o cálculo explícito de uma dessas potências, como A^{10} por exemplo, a alternativa seguinte utiliza apenas o cálculo do polinômio característico $p_A(\lambda)$, o que é bem mais factível: Usa-se o algoritmo da divisão para escrever $\lambda^k = p_A(\lambda) \cdot q(\lambda) + r(\lambda)$, onde o grau do resto $r(\lambda)$ é menor do que a dimensão do espaço E. Tem-se então $A^k = r(A)$. Usando este método, mostre que, dado o operador $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, onde $A(x, y, z) = (x+2y+3z, 3x+y+2z, x-y+z)$, tem-se $A^5 = 40A^2 + 19A - 143 \cdot I$ e $A^{10} = 14281 \cdot A^2 + 2246 \cdot A - 49071 \cdot I$.

22.12. Para cada uma das equações abaixo, determine a solução que tem os valores iniciais indicados.

(a) $x_{k+2} = x_{k+1} + 20x_k$, $x_o = -3$, $x_1 = 2$.

(b) $\frac{1}{5}x_{k+2} = 2x_{k+1} - 5x_k$, $x_o = 2$, $x_1 = -1$.

(c) $x_{k+2} - 6x_{k+1} + 25x_k = 0$, $x_o = 1$, $x_2 = 3$.

22.13. A *seqüência de Fibonacci* $(x_o, x_1, \ldots, x_k, \ldots)$ é definida pelas condições $x_o = 0$, $x_1 = 1$ e $x_{k+2} = x_{k+1} + x_k$. Obtenha a fórmula geral para x_k em função de k, prove que $x_{k+2} = 1 + x_1 + \cdots + x_k$ e que

$$\lim_{k\to\infty} \frac{x_{k+1}}{x_k} = \frac{1+\sqrt{5}}{2} \quad \text{(o } \textit{número de ouro}\text{)}.$$

22.14. Qual a fórmula que exprime x_k em função de x_o, x_1 e k, sabendo-se que

$$x_{k+2} = \frac{1}{2}(x_{k+1} + x_k)$$

para todo $k \geq 0$?

22.15. Resolva a equação $x_{k+3} - 6x_{k+2} + 11x_{k+1} - 6x_k = 0$.

22.16. Ache a solução $v_k = (x_k, y_k)$ do sistema

$$\begin{aligned} x_{k+1} &= x_k - \alpha(x_k - y_k) \\ y_{k+1} &= y_k + \beta(x_k - y_k), \end{aligned}$$

com vetor inicial $v_o = (x_o, y_o)$, com $y_o < x_o$, $0 < \alpha < 1$ e $0 < \beta < 1$. Mostre que $\lim x_k = \lim y_k = (\beta x_o + \alpha y_o)/(\alpha + \beta)$, logo este limite está mais próximo de x_o do que de y_o se, e somente se, $\alpha < \beta$. Mostre que se tem $x_k < y_k$ para todo k se, e somente se, $\alpha + \beta < 1$, em cujo caso a seqüência (x_k) é decrescente e (y_k) é crescente.

[**Observação:** este sistema é um modelo para uma situação simples de barganha. Cada x_k é o preço do vendedor e y_k é a proposta do comprador. Em cada etapa, o vendedor oferece um desconto proporcional à diferença de preços na etapa anterior e o comprador, por sua vez, aumenta sua proposta de modo análogo. Se a soma $\alpha + \beta$, da constante do vendedor com a do comprador, for maior do que 1, já na primeira etapa tem-se $x_1 < y_1$, o que daria o chamado “negócio de pai para filho”...]

22.17. Seja $x_{k+2} + a x_{k+1} + b x_k = 0$ uma equação cujas raízes características são os complexos conjugados r e $\overline{r}$. Escreva $r = \rho(\cos\theta + i \operatorname{sen}\theta)$ como $x_k = \alpha\rho^k \cos(\beta + k\theta)$, onde as constantes α e β podem ser determinadas de modo a fazer com que x_o e x_1 assumam os valores iniciais pré-estabelecidos. [Sugestão: a equação dada admite a solução geral complexa $x_k = \zeta r^k + \eta \overline{r}^k$, onde $\zeta, \eta \in \mathbb{C}$ são arbitrários. Tomando $\eta = \overline{\zeta}$, obtém-se a solução real $x_k = 2\,\mathrm{Re}\,(\zeta r^k)$. Escreva $\zeta = \frac{\alpha}{2}(\cos\beta + i \operatorname{sen}\beta)$ e use a fórmula $\cos(x + y) = \cos x \cdot \cos y - \operatorname{sen} x \cdot \operatorname{sen} y$.]

22.18. Dada a equação $x_{k+2} + a x_{k+1} + b x_k = c^k$, onde c não é raiz característica, prove que existe $M \in \mathbb{R}$ tal que $x_k = M \cdot c^k$ é uma solução particular. Se c é uma raiz característica diferente de $-a/2$, prove que existe N tal que $x_k = Nkc^k$ é uma solução particular. E se $c = -a/2$ é raiz característica prove que existe P tal que $x_k = Pk^2c^k$ é uma solução particular.

22.19. Se $v_{k+1} = \alpha v_k + w_k$, onde $|\alpha| < 1$ e $\lim_{k\to\infty} w_k = 0$, prove que $\lim_{k\to\infty} v_k = 0$

22.20. Seja $A \colon E \to E$ um operador linear cujos autovalores cumprem $|\lambda_1| < 1, \ldots, |\lambda_n| < 1$. Prove que, para todo $v \in E$, tem-se $\lim_{k\to\infty} A^k v = 0$. [Sugestão: Tome uma base $\{u_1, \ldots, u_n\} \subset E$ na qual a matriz de A seja diagonal superior. Observe que $Au_i = w + \lambda_i u_i$, onde $w \in S(u_1, \ldots, u_{i-1})$ se $i > 1$ e $Au_1 = \lambda_1 u_1$. Logo $A^{k+1} u_i = A^k w + \lambda_i A^k u_i$. Ponha $v_k = A^k u_i$, $w_k = A^k w$, $\alpha = \lambda_i$ e tenha $v_{k+1} = \alpha v_k + w_k$. Use indução em i e o exercício anterior para concluir que $\lim_{k\to\infty} A^k u_i = 0$ $(i = 1, \ldots, n)$, e daí $\lim_{k\to\infty} A^k v = 0$ para todo $v \in E$.]

Apêndice

A Forma Canônica de Jordan

O objetivo deste apêndice é provar que, dado um operador linear $A: E \to E$ *num espaço vetorial complexo de dimensão finita, existe uma base de* E *na qual a matriz* **a** *de* A *é formada por uma série de "blocos de Jordan" ao longo da diagonal. Um bloco de Jordan é uma matriz triangular inferior cujos elementos diagonais são todos iguais a um mesmo auto-valor de* A *e os elementos imediatamente abaixo da diagonal são iguais a 1. Diz-se então que a matriz* **a** *está na forma canônica de Jordan. Quando* E *possui uma base formada por auto-vetores de* A*, os blocos de Jordan são todos* 1×1*, e neste caso a forma canônica de Jordan para* A *é uma matriz diagonal.*

A forma canônica de Jordan exibe a matriz mais simples que se pode obter para o operador A*. Ela se mostra útil no estudo de questões que envolvem potências sucessivas do operador* A*, como as equações diferenciais lineares e as equações a diferenças finitas lineares.*

A1. Operadores Nilpotentes

Nesta seção, estudaremos mais um tipo de operadores que podem ser representados por matrizes especialmente simples, a saber, os

operadores nilpotentes. Os espaços vetoriais aqui considerados podem ser reais ou complexos; não faz diferença. Tampouco se necessita de um produto interno. Veremos que, mesmo no caso real, os operadores nilpotentes possuem matrizes triangulares. O estudo aqui feito serve de preparação para a seção seguinte.

Um operador linear $A\colon E \to E$ diz-se *nilpotente* quando se tem $A^k = 0$ para algum $k \in \mathbb{N}$. O *índice* de um operador nilpotente é o menor número $k \in \mathbb{N}$ tal que $A^k = 0$. Isto significa que $A^{k-1} \neq 0$ e $A^k = 0$.

Analogamente, uma matriz quadrada **a** chama-se *nilpotente* quando se tem $\mathbf{a}^k = 0$ para algum $k \in \mathbb{N}$. Se $\mathbf{a}^{k-1} \neq 0$ e $\mathbf{a}^k = 0$, diz-se que a matriz nilpotente **a** tem *índice* k.

Exemplo A1.1. O operador de derivação $D\colon \mathcal{P}_n \to \mathcal{P}_n$ é nilpotente, com índice $n+1$.

Exemplo A1.2. Um exemplo simples de matriz nilpotente é dado pela matriz $k \times k$ cuja k-ésima coluna é o vetor nulo e, para $1 \leq j \leq k-1$, sua j-ésima coluna é $e_{j+1} \in \mathbb{R}^k$. Para $k = 4$ essa matriz tem a forma abaixo:

$$\mathbf{a} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

A matriz deste exemplo provém do operador $A\colon \mathbb{R}^k \to \mathbb{R}^k$, definido por $Ae_1 = e_2, \ldots, Ae_{k-1} = e_k$, $Ae_k = 0$. Evidentemente, tem-se $A^k = 0$ e $A^{k-1} \neq 0$. Logo o índice do operador A (e da matriz **a**) é igual a k.

Teorema A1.1. *Dado o operador* $A\colon E \to E$*, seja* $u \in E$ *um vetor tal que* $A^{k-1}u \neq 0$ *e* $A^k u = 0$*. Então os vetores* $u, Au, \ldots, A^{k-1}u$ *são linearmente independentes.*

Demonstração: Seja

$$\alpha_1 u + \alpha_2 Au + \cdots + \alpha_k A^{k-1} u = 0.$$

Aplicando o operador A^{k-1} a ambos os membros desta igualdade, obtemos $\alpha_1 A^{k-1} u = 0$. Como $A^{k-1} u \neq 0$, concluímos que $\alpha_1 = 0$. Logo a combinação linear inicial se reduz a

$$\alpha_2 Au + \cdots + \alpha_k A^{k-1} u = 0.$$

Aplicando o operador A^{k-2}, obtemos agora $\alpha_2 A^{k-1} u = 0$, logo $\alpha_2 = 0$. Prosseguindo analogamente, tem-se $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$. □

Corolário 1. *Num espaço vetorial de dimensão n, o índice de um operador nilpotente é $\leq n$.*

Corolário 2. *Seja $A \colon E \to E$ um operador nilpotente de índice n num espaço vetorial E, de dimensão n. Existe uma base de E na qual a matriz de A tem a forma abaixo:*

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Vale, evidentemente, a recíproca do Corolário 2 acima: se alguma matriz do operador $A \colon E \to E$ (onde $\dim E = n$) tem a forma acima então A é um operador nilpotente de índice n.

Se o índice do operador nilpotente $A \colon E \to E$ for menor do que a dimensão do espaço E, mostraremos a seguir que existe uma base de E na qual a matriz de A é formada por blocos do tipo acima, dispostos ao longo da diagonal.

A idéia da demonstração é extremamente simples, mas a notação pode tornar-se longa. A fim de evitar complicações tipográficas, trataremos os casos de índices mais baixos, deixando claro o processo indutivo que leva ao caso geral.

O argumento se baseia no seguinte fato, que foi estabelecido na demonstração do Teorema do Núcleo e da Imagem, e que destacaremos aqui como um lema:

Lema. *Se $\{Au_1, \ldots, Au_p\}$ é uma base da imagem do operador $A \colon E \to E$ e $\{v_1, \ldots, v_q\}$ é uma base do núcleo de A então $\{u_1, \ldots, u_p, v_1, \ldots, v_q\}$ é uma base de E.*

Seja inicialmente o operador nilpotente $A \colon E \to E$, de índice 2: $A \neq 0$ e $A^2 = 0$.

Tomemos uma base $\{Au_1, \ldots, Au_p\}$ da imagem de A. A condição $A^2 = 0$ significa que $\mathcal{I}m(A) \subset \mathcal{N}(A)$, logo existem vetores $v_1, \ldots, v_q$ tais que

$$\mathcal{U} = \{Au_1, \ldots, Au_p, v_1, \ldots, v_q\}$$

é uma base de $\mathcal{N}(A)$. Pelo Lema, o conjunto

$$\mathcal{V} = \{u_1, Au_1, \ldots, u_p, Au_p, v_1, \ldots, v_q\}$$

é uma base de E.

Em relação a esta base $\mathcal{V}$, a matriz do operador nilpotente $A\colon E \to E$, de índice 2, é formada por p blocos de matrizes 2×2 do tipo

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

ao longo da diagonal (onde p é o posto de A), seguidos de q colunas nulas, onde $2p + q = \dim E$.

Por exemplo, se $A\colon \mathbb{R}^5 \to \mathbb{R}^5$ é nilpotente de índice 2, sua matriz na base $\mathcal{V}$ tem uma das formas abaixo, conforme seu posto seja 2 ou 1:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Em seguida, consideremos um operador nilpotente $A\colon E \to E$ de índice 3.

A restrição de A ao subespaço invariante $\mathcal{I}m(A)$ é um operador nilpotente de índice 2. Lembrando que os elementos de $\mathcal{I}m(A)$ são todos da forma Au, resulta do que vimos acima que existe uma base de $\mathcal{I}m(A)$ do tipo

$$\{Au_1, A^2u_1, \ldots, Au_p, A^2u_p, Av_1, \ldots, Av_q\},$$

com $A^2v_1 = \cdots = A^2v_q = 0$. Os vetores linearmente independentes $A^2u_1, \ldots, A^2u_p, Av_1, \ldots, Av_q$ pertencem ao núcleo de A, logo podem ser incluídos numa base:

$$\mathcal{U} = \{A^2u_1, \ldots, A^2u_p, Av_1, \ldots, Av_q, w_1, \ldots, w_r\} \subset \mathcal{N}(A).$$

Segue do Lema que o conjunto

$$\mathcal{V} = \{u_1, Au_1, A^2u_1, \ldots, \\ u_p, Au_p, A^2u_p, v_1, Av_1, \ldots, v_q, Av_q, w_1, \ldots, w_r\}$$

é uma base de E.

Em relação a esta base $\mathcal{V}$, a matriz do operador nilpotente $A\colon E \to E$, de índice 3, é formada por p blocos de matrizes 3×3 da forma

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

ao longo da diagonal, seguidos por q blocos de matrizes 2×2 da forma

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

ainda ao longo da diagonal, e por r colunas de zeros. (Aqui, p é o posto de A^2, $2p+q$ é o posto de A e $p+q+r$ é a dimensão de $\mathcal{N}(A)$.)

Eventualmente, pode-se ter $q = 0$ ou $r = 0$ (ou ambos). Mas as três primeiras colunas do operador nilpotente A, de índice 3, na base $\mathcal{V}$, devem ser e_2, e_3 e 0.

A discussão acima assegura, para um operador nilpotente $A\colon \mathbb{R}^5 \to \mathbb{R}^5$ de índice 3, uma base $\mathcal{V}$ na qual sua matriz tem uma das formas seguintes

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

conforme o posto de A seja 3 ou 2.

O caso geral se trata da mesma maneira. A fim de dar mais precisão e clareza ao seu enunciado, vamos introduzir uma definição.

Dado um operador nilpotente $A\colon E \to E$, dizemos que um subespaço vetorial $F \subset E$ é *cíclico* (em relação a A) quando existe um vetor $u \in F$ tal que $A^m u = 0$ e $\{u, Au, \ldots, A^{m-1}u\}$ é uma base de F. Isto significa que $F \subset E$ é um subespaço vetorial de dimensão m, invariante por A, e que a restrição de A ao subespaço F é um operador nilpotente de índice m.

Por exemplo, na base $\mathcal{V}$, acima obtida quando analisamos um operador nilpotente de índice 3, cada um dos vetores $u_1, \ldots, u_p$ gera um subespaço cíclico de dimensão 3, cada $v_j (j = 1, \ldots, q)$ gera um

subespaço cíclico de dimensão 2 e cada $w_\ell (\ell = 1, \ldots, r)$ gera um subespaço cíclico de dimensão 1 (o que significa $Aw_1 = \cdots = Aw_\ell = 0$).

O resultado fundamental sobre operadores nilpotentes é o

Teorema A1.2. *Seja* $A \colon E \to E$ *um operador nilpotente de índice* k *num espaço vetorial de dimensão* n. *Existem inteiros* $k_1 = k \geq k_2 \geq \cdots \geq k_r > 0$, *tais que* $E = F_1 \oplus \cdots \oplus F_r$, *onde cada* F_i *é um subespaço cíclico de dimensão* k_i.

Evidentemente, $k_1 + \cdots + k_r = n$.

Tomando em cada F_i $(i = 1, \ldots, r)$ uma base $\mathcal{V}_i = \{u_i, Au_i, \ldots, A^{k_i - 1} u_i\}$, obtemos uma base $\mathcal{V} = \mathcal{V}_1 \cup \ldots \cup \mathcal{V}_r$, em relação à qual a matriz de A é formada por r blocos $\mathbf{a}_i \in M(k_i \times k_i)$, ao longo da diagonal. Cada bloco $\mathbf{a}_i$ tem a forma vista no Exemplo 2: para $j < k_i$ sua j-ésima coluna é $e_{j+1} \in \mathbb{R}^{k_i}$ enquanto sua k_i-ésima coluna é zero.

A2. Existência da Forma Canônica de Jordan.

Dado um operador linear $A \colon E \to E$ num espaço vetorial complexo de dimensão finita, provaremos que existe uma base em E na qual a matriz de A tem a forma canônica de Jordan: é triangular inferior, os auto-valores que formam sua diagonal são repetidos consecutivamente de acordo com suas multiplicidades algébricas e, além disso, os elementos imediatamente abaixo da diagonal são iguais a 0 ou 1; todos os demais elementos são nulos.

Teorema A2.1. *Seja* $A \colon E \to E$ *um operador linear num espaço vetorial (real ou complexo) de dimensão finita. Existe uma decomposição* $E = F \oplus G$, *como soma direta de subespaços invariantes* F, G *tais que* A *é nilpotente em* F *e invertível em* G.

Demonstração: Como a dimensão de E é finita, a seqüência de subespaços invariantes

$$E \supset \mathcal{I}m(A) \supset \mathcal{I}m(A^2) \supset \ldots$$

não pode ser estritamente decrescente para sempre. Seja então k o menor número natural tal que $\mathcal{I}m(A^k) = \mathcal{I}m(A^{k+1})$. Afirmamos que então $\mathcal{I}m(A^{k+1}) = \mathcal{I}m(A^{k+2})$. Com efeito,

$$\mathcal{I}m(A^{k+2}) = A[\mathcal{I}m(A^{k+1})] = A[\mathcal{I}m(A^k)] = \mathcal{I}m(A^{k+1}).$$

Segue-se que $\mathcal{I}m(A^{k+2}) = \mathcal{I}m(A^{k+3})$, etc. Note-se que vale

$$\mathcal{N}(A) \subset \mathcal{N}(A^2) \subset \cdots \subset \mathcal{N}(A^k) = \mathcal{N}(A^{k+1}) = \mathcal{N}(A^{k+2}) = \cdots .$$

Com efeito, pelo Teorema do Núcleo e da Imagem, temos

$$\begin{aligned}\dim \mathcal{N}(A^{k+1}) &= \dim E - \dim \mathcal{I}m(A^{k+1}) \\ &= \dim E - \dim \mathcal{I}m(A^k) = \dim \mathcal{N}(A^k).\end{aligned}$$

Sejam $F = \mathcal{N}(A^k)$ e $G = \mathcal{I}m(A^k)$. Evidentemente, F e G são invariantes por A e a restrição $A\colon F \to F$ é nilpotente. Além disso, a restrição $A\colon G \to G$ é um operador sobrejetivo pois

$$A(G) = A[\mathcal{I}m(A^k)] = \mathcal{I}m(A^{k+1}) = \mathcal{I}m(A^k) = G.$$

Logo $A\colon G \to G$ é invertível. Mostremos agora que $E = F + G$. Dado $v \in E$, como $\mathcal{I}m(A^k) = \mathcal{I}m(A^{2k})$, existe $x \in E$ tal que $A^k v = A^{2k} x$. Então, se escrevermos

$$v = (v - A^k x) + A^k x,$$

veremos que $A^k(v - A^k x) = A^k v - A^{2k} x = 0$, logo $v - A^k x \in F$ e, obviamente, $A^k x \in G$. Assim, todo elemento $v \in E$ é soma de um vetor de F com um vetor de G, ou seja, $E = F + G$. Para concluir que $E = F \oplus G$, resta apenas mostrar que $F \cap G = \{0\}$. Ora, sabemos que

$$\begin{aligned}\dim F + \dim G &= \dim(F + G) + \dim(F \cap G) \\ &= \dim E + \dim(F \cap G).\end{aligned}$$

Por outro lado, o Teorema do Núcleo e da Imagem, aplicado ao operador $A^k\colon E \to E$, nos dá $\dim E = \dim F + \dim G$. Segue-se então que $\dim(F \cap G) = 0$, isto é, $F \cap G = \{0\}$. □

Teorema A2.2. *Seja* $E = F \oplus G$ *como no Teorema A2.1. Se* n_0 *é a multiplicidade algébrica do autovalor 0 do operador* $A\colon E \to E$ *então a dimensão do subespaço* F *é igual a* n_0*. Além disso,* F *é o núcleo e* G *é a imagem de* $A^{n_0}\colon E \to E$*. Segue-se daí que a decomposição* $E = F \oplus G$*, com as propriedades enunciadas naquele teorema, é única.*

Demonstração: Sejam $A'\colon F \to F$ e $A''\colon G \to G$ as restrições do operador A aos subespaços invariantes F e G. Como A' é nilpotente e A'' é invertível, o polinômio característico de A' é $p_{A'}(\lambda) = (-\lambda)^n$,

$n = \dim F$, e o de A'' cumpre a condição $p_{A''}(0) \neq 0$. A prova do lema que antecede o Teorema 20.1 nos dá $p_A = p_{A'} \cdot p_{A''}$. Por outro lado, $p_A(\lambda) = \lambda^{n_0} \cdot q(\lambda)$, com $q(0) \neq 0$. Assim, $\lambda^n \cdot p_{A''}(\lambda) = \lambda^{n_0} \cdot q(\lambda)$, com $p_{A''}(0) \neq 0$ e $q(0) \neq 0$. Segue-se que $n = n_0$. Sendo A nilpotente no subespaço F de dimensão n_0, tem-se $F \subset \mathcal{N}(A^{n_0})$. Reciprocamente, se $u \in \mathcal{N}(A^{n_0})$, escrevemos $u = v + w$, com $v \in F$ (logo $A^{n_0}v = 0$) e $w \in G$. Então $0 = A^{n_0}v + A^{n_0}w = A^{n_0}w$. Sendo A invertível em G, de $A^{n_0}w = 0$ conclui-se que $w = 0$, logo $u = v \in F$. Assim, $F = \mathcal{N}(A^{n_0})$. Para provar que $G = \mathcal{I}m(A^{n_0})$, observamos primeiro que, sendo A invertível em G, o operador $A^{n_0} \colon G \to G$ também é invertível, logo $G \subset \mathcal{I}m(A^{n_0})$. Por outro lado, para todo $u \in E$, escrevendo $u = v + w$ com $v \in F$ e $w \in G$, temos $A^{n_0}u = A^{n_0}w \in G$ (pois G é invariante por A) logo $\mathcal{I}m(A^{n_0}) \subset G$. Assim, $\mathcal{I}m(A^{n_0}) = G$. □

Observação. Para uso na demonstração do próximo teorema, notemos aqui que se $E = F_1 + \cdots + F_r$ e $\dim E \geq \dim F_1 + \cdots + \dim F_r$ então $E = F_1 \oplus \cdots \oplus F_r$. Com efeito, tomando em cada subespaço F_i uma base $\mathcal{V}_i$ $(i = 1, \ldots, r)$, o conjunto $\mathcal{V} = \mathcal{V}_1 \cup \ldots \cup \mathcal{V}_r$ gera E e o número de elementos de $\mathcal{V}$ é $\leq \dim E$, logo $\mathcal{V}$ é uma base de E. Assim, todo vetor $v \in E$ se exprime, de modo único, como soma $v = v_1 + \cdots + v_r$, com $v_1 \in F_1, \ldots, v_r \in F_r$. Noutras palavras, $E = F_1 \oplus \cdots \oplus F_1$.

Teorema A2.3. *Sejam* $\lambda_1, \ldots, \lambda_r$ *os auto-valores distintos do operador* $A \colon E \to E$*, num espaço vetorial complexo de dimensão finita. Para cada* $i = 1, \ldots, r$*, sejam* n_i *a multiplicidade algébrica de* λ_i *e* $E_i = \mathcal{N}[(A - \lambda_i I)^{n_i}]$*. Então* $\dim E_i = n_i$ *e* $E = E_1 \oplus \cdots \oplus E_r$.

Demonstração: Mostremos inicialmente que n_i é também a multiplicidade algébrica do auto-valor 0 do operador $A_i = A - \lambda_i I$. Com efeito, $p_{A_i}(\lambda) = \det[(A - \lambda_i I) - \lambda I] = \det[A - (\lambda + \lambda_i)I] = p_A(\lambda + \lambda_i)$. Temos $p_A(\lambda) = (\lambda - \lambda_i)^{n_i} q(\lambda)$ com $q(\lambda_i) \neq 0$. Logo $p_{A_i}(\lambda) = p_A(\lambda + \lambda_i) = \lambda^{n_i} \cdot r(\lambda)$, onde $r(\lambda) = q(\lambda + \lambda_i)$, portanto $r(0) \neq 0$. Isto posto, o Teorema A2.2 nos assegura que $\dim E_i = n_i$. Em particular, $\dim E_1 + \cdots + \dim E_r = \dim E$. Pela observação que precede este teorema, resta-nos apenas provar que $E = E_1 + \cdots + E_r$. Ora, o polinômio característico do operador A se decompõe na forma

$$p_A(\lambda) = \prod_{j=1}^{r} (\lambda - \lambda_j)^{n_j}.$$

Se pusermos

$$q_i(\lambda) = \prod_{j \neq i} (\lambda - \lambda_j)^{n_j},$$

obteremos os polinômios $q_1(\lambda), \ldots, q_r(\lambda)$, primos entre si. Por um conhecido teorema de Álgebra, existem polinômios $m_1(\lambda), \ldots, m_r(\lambda)$ tais que

$$m_1(\lambda)q_1(\lambda) + \cdots + m_r(\lambda)q_r(\lambda) = 1.$$

Segue-se que

$$m_1(A)q_1(A) + \cdots + m_r(A)q_r(A) = I.$$

Assim, para todo $v \in E$, tem-se

$$v = v_1 + \cdots + v_r\,, \quad v_i = m_i(A)q_i(A)v.$$

Pelo Teorema de Cayley-Hamilton, temos

$$\begin{aligned} A_i^{n_i} \cdot q_i(A) &= (A - \lambda_i I)^{n_i} \cdot \prod_{j \neq i} (A - \lambda_j I)^{n_j} \\ &= \prod_{j=1}^{r} (A - \lambda_j I)^{n_j} = p_A(A) = 0. \end{aligned}$$

Logo $A_i^{n_i} v_i = 0$, ou seja $v_i \in E_i$ para todo $i = 1, \ldots, r$. Isto conclui a demonstração do teorema. □

Teorema A2.4. *Os subespaços* $E_i = \mathcal{N}[(A - \lambda_i I)^{n_i}]$ *definidos no Teorema A2.3 são invariantes por qualquer operador* $B\colon E \to E$ *que comute com* A.

Demonstração: $AB = BA \Rightarrow (A - \lambda_i I)B = B(A - \lambda_i I) \Rightarrow$ $\Rightarrow (A - \lambda_i I)^{n_i} B = B(A - \lambda_i I)^{n_i}$. Logo $v \in E_i \Rightarrow (A - \lambda_i I)^{n_i} Bv = B(A - \lambda_i I)^{n_i} v = B \cdot 0 = 0 \Rightarrow Bv \in E_i$. □

Corolário. *Os subespaços* $E_1, \ldots, E_r$ *são invariantes por* A.

Um *bloco de Jordan* $n \times n$ é uma matriz triangular inferior da forma

$$B(\lambda; n) = \begin{bmatrix} \lambda & & & & \\ 1 & \lambda & & & \\ & 1 & \ddots & & \\ & & \ddots & \lambda & \\ & & & 1 & \lambda \end{bmatrix}$$

onde os elementos da diagonal são todos iguais, os elementos imediatamente abaixo da diagonal são todos iguais a 1 e os demais elementos são zeros.

Diz-se que uma matriz está na *forma canônica de Jordan* quando ela é triangular inferior, com blocos de Jordan ao longo da diagonal e os demais elementos iguais a zero. Os blocos de Jordan devem estar agrupados consecutivamente em listas do tipo

$$B(\lambda_i; k_1), B(\lambda_i; k_2), \ldots, B(\lambda_i; k_{s_i}),$$

onde $k_1 + k_2 + \cdots + k_{s_i} = n_i =$ multiplicidade algébrica do auto-valor λ_i da matriz dada.

Por exemplo, dispondo os blocos $B(\lambda_1; 3)$, $B(\lambda_1; 1)$ e $B(\lambda_2; 2)$ ao longo da diagonal, obtemos uma matriz 6×6 na forma canônica de Jordan:

$$\begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 1 & \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_2 \end{bmatrix}$$

Teorema A2.5. *Para todo operador* $A \colon E \to E$ *num espaço vetorial complexo de dimensão finita, existe uma base na qual a matriz de* A *tem a forma canônica de Jordan.*

Demonstração: Seja $E = E_1 \oplus \cdots \oplus E_r$ a decomposição assegurada pelo Teorema A2.3. O Teorema A1.2 prova a existência da forma canônica de Jordan para operadores nilpotentes. Ora, para cada $i = 1, \ldots, r$, a restrição $A - \lambda_i I \colon E_i \to E_i$ é nilpotente. Logo existe uma base $\mathcal{V}_i \subset E_i$ na qual a matriz de $A - \lambda_i I \colon E_i \to E_i$ tem a forma canônica de Jordan (com zeros na diagonal). Logo a matriz da restrição $A = (A - \lambda_i I) + \lambda_i I \colon E_i \to E_i$ tem a forma canônica de Jordan (com os elementos da diagonal todos iguais a λ_i). Segue-se que $\mathcal{V} = \mathcal{V}_1 \cup \ldots \cup \mathcal{V}_r$ é uma base de E na qual a matriz de A tem a forma canônica de Jordan. □

Do ponto de vista matricial, o resultado que acabamos de provar significa que, para toda matriz quadrada complexa $\mathbf{a}$, existe uma matriz (complexa) invertível $\mathbf{p}$ tal que $\mathbf{p}^{-1}\mathbf{a}\mathbf{p}$ está na forma canônica de Jordan.

Exemplo A2.1. Vamos usar a forma canônica de Jordan para provar que toda matriz invertível possui uma raiz quadrada (complexa). Preliminarmente observamos que se $\mathbf{x}$ é uma raiz quadrada de $\mathbf{p}^{-1}\mathbf{ap}$ então $\mathbf{pxp}^{-1}$ é uma raiz quadrada de $\mathbf{a}$, pois

$$(\mathbf{pxp}^{-1})^2 = \mathbf{pxp}^{-1}\mathbf{pxp}^{-1} = \mathbf{px}^2\mathbf{p}^{-1} = \mathbf{p}(\mathbf{p}^{-1}\mathbf{ap})\mathbf{p}^{-1} = \mathbf{a}.$$

Portanto, ao provar a existência da raiz quadrada de uma matriz invertível, não há perda de generalidade em supor que essa matriz está na forma canônica de Jordan, que é uma forma triangular particular. Em virtude da invertibilidade, os elementos da diagonal (auto-valores) são todos diferentes de zero.

Trataremos explicitamente do caso 4×4, deixando para o leitor o caso 3×3 (mais simples) e o caso geral (mais complicado, porém suscetível da mesma abordagem).

Temos então uma matriz da forma

$$\mathbf{a} = \begin{bmatrix} a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ 0 & d & e & 0 \\ 0 & 0 & f & g \end{bmatrix},$$

com a, c, e, g diferentes de zero, e procuramos uma matriz

$$\mathbf{x} = \begin{bmatrix} x & 0 & 0 & 0 \\ y & z & 0 & 0 \\ m & n & p & 0 \\ q & r & s & t \end{bmatrix}$$

tal que $\mathbf{x}^2 = \mathbf{a}$. Ora, um cálculo simples nos dá

$$\mathbf{x}^2 = \begin{bmatrix} x^2 & 0 & 0 & 0 \\ y(x+z) & z^2 & 0 & 0 \\ m(x+p)+ny & n(z+p) & p^2 & 0 \\ q(x+t)+ry+ms & r(z+t)+ns & s(p+t) & t^2 \end{bmatrix}.$$

Portanto as incógnitas x, y, z, m, n, p, q, r, s e t devem satisfazer as condições

(1) $\quad x^2 = a, \quad z^2 = c, \quad p^2 = e, \quad t^2 = g,$

(2) $\quad y(x+z) = b, \quad n(z+p) = d, \quad s(p+t) = f,$

(3) $\quad m(x+p)+ny=0, \quad r(z+t)+ns=0,$

(4) $\quad q(x+t)+ry+ms=0.$

Como a, c, e, g são diferentes de zero, podemos escolher x, z, p, t tais que as igualdades (1) sejam satisfeitas e, além disso, se tenha $x+z\neq 0$, $z+p\neq 0$, $p+t\neq 0$, $x+p\neq 0$, $z+t\neq 0$ e $x+t\neq 0$. Isto nos permite determinar y, n, s de modo a satisfazer as igualdades (2), em seguida obter m, r de modo que as igualdades (3) sejam válidas e, finalmente, usar (4) para determinar q.

Observação: Uma matriz não-invertível pode não possuir raiz quadrada. Este é o caso, por exemplo, da matriz

$$\mathbf{a}=\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}.$$

É fácil ver que não existem números complexos x, y, z, t tais que a matriz $\mathbf{x}=\begin{bmatrix}x & y\\ z & t\end{bmatrix}$ cumpra $\mathbf{x}^2=\mathbf{a}$.

A3. A Decomposição A = N+D

Nesta seção, mostraremos como visualizar a forma canônica de Jordan de modo intrínseco, exprimindo-a sob o ponto-de-vista de operadores, em vez de matrizes.

A forma canônica de Jordan, estabelecida na seção anterior, mostra que, dado um operador $A\colon E\to E$ num espaço vetorial complexo de dimensão finita, existe uma base $\mathcal{V}\subset E$ na qual a matriz $\mathbf{a}$ de A é formada por blocos de Jordan ao longo da diagonal, sendo os blocos que correspondem ao mesmo auto-valor de A agrupados consecutivamente. Segue-se que $\mathbf{a}=\mathbf{n}+\mathbf{d}$, onde $\mathbf{d}$ é uma matriz diagonal, os elementos dessa diagonal sendo os auto-valores de A repetidos de acordo com sua multiplicidade, e $\mathbf{n}$ é uma matriz triangular inferior nilpotente (logo os elementos de sua diagonal são todos iguais a zero) na qual os elementos imediatamente abaixo da diagonal são iguais a 1 ou a 0 e os demais elementos são nulos.

Resulta imediatamente daí a decomposição $A=N+D$, onde $D\colon E\to E$ é o operador cuja matriz na base $\mathcal{V}$ é $\mathbf{d}$ e $N\colon E\to E$ é o operador nilpotente do qual $\mathbf{n}$ é a matriz na mesma base $\mathcal{V}$.

Um operador $D\colon E \to E$ chama-se *diagonalizável* quando existe alguma base de E na qual a matriz de D é diagonal. Isto equivale a dizer que a referida base de E é formada por auto-vetores do operador D.

Assim, acabamos de mostrar que, num espaço vetorial complexo de dimensão finita, todo operador $A\colon E \to E$ pode escrever-se como soma $A = N + D$ de um operador nilpotente com um diagonalizável.

Na notação do Teorema A2.3, N é o operador cuja restrição a cada subespaço E_i coincide com $A - \lambda_i I$, enquanto D restrito a cada um dos E_i's é igual a $\lambda_i I$. Como $A - \lambda_i I$ e $\lambda_i I$ comutam para todo $i = 1, \ldots, r$, segue-se que $ND = DN$.

Provaremos a seguir que esta é a única maneira de se escrever $A = N + D$ com N nilpotente, D diagonalizável e $ND = DN$.

Para maior clareza, destacaremos sob a forma de lemas dois fatos elementares que usaremos na demonstração dessa unicidade.

Lema 1. *A restrição de um operador diagonalizável* $D\colon E \to E$ *a um subespaço invariante* $F \subset E$ *é ainda um operador diagonalizável* $D\colon F \to F$.

Demonstração: Seja $\mathcal{V} \subset E$ uma base formada por auto-vetores de D. Introduzimos em E um produto interno hermitiano, impondo que a base $\mathcal{V}$ seja ortonormal. Relativamente a esse produto interno, D é normal. Portanto a restrição $D\colon F \to F$ é um operador hermitiano, logo diagonalizável. (V. Exercício 15.19). □

Lema 2. *A soma de dois operadores nilpotentes que comutam é ainda um operador nilpotente.*

Demonstração: Sejam $M, N\colon E \to E$ com $M^p = 0$, $N^q = 0$ e $MN = NM$. Esta comutatividade assegura que vale o binômio de Newton:

$$(M+N)^{p+q} = \sum_{i=0}^{p+q} \binom{p+q}{i} M^i N^{p+q-i}.$$

No somatório acima, as parcelas com $i \geq p$ são nulas porque, neste caso, $M^i = 0$. Se, entretanto, tem-se $i < p$ então $p + q - i > q$, logo $N^{p+q-i} = 0$. Assim as parcelas com $i < p$ também são nulas e concluímos que $(M + N)^{p+q} = 0$. □

Teorema A3.1. *Seja* E *um espaço vetorial complexo de dimensão finita. Para todo operador linear* $A \colon E \to E$*, existe uma única decomposição* $A = N + D$ *com* $N \colon E \to E$ *nilpotente,* $D \colon E \to E$ *diagonalizável e* $ND = DN$.

Demonstração: Evidentemente, N e D comutam com A. Pelo Teorema A2.4, cada subespaço $E_i = \mathcal{N}[(A - \lambda_i I)^{n_i}]$ é invariante por N e por D. Para $i = 1, \ldots, r$, sejam $A_i, N_i, D_i \colon E_i \to E_i$ as restrições de A, N e D ao subespaço E_i. A igualdade $A_i = N_i + D_i$ pode ser escrita como $(A_i - \lambda_i I) + \lambda_i I = N_i + D_i$ ou, ainda, como

$$(A_i - \lambda_i I) - N_i = D_i - \lambda_i I. \qquad (*)$$

Pelo Lema 2, o operador $(A_i - \lambda_i I) - N_i$ é nilpotente e pelo Lema 1, D_i é diagonalizável, logo $D_i - \lambda_i I$ é diagonalizável (pois qualquer vetor não-nulo é auto-vetor de $\lambda_i I$). Pela igualdade (*), esses operadores são, ao mesmo tempo, nilpotentes e diagonalizáveis, logo iguais a zero. Portanto vale $N_i = A_i - \lambda_i I$ e $D_i = \lambda_i I$ para $i = 1, \ldots, r$. Segue-se que N e D são os operadores anteriormente obtidos a partir do Teorema A2.3. □

Indicações Bibliográficas

Na seqüência usual das disciplinas matemáticas que se estudam na universidade, a Álgebra Linear é pré-requisito para a Análise das funções de várias variáveis e para as Equações Diferenciais Ordinárias. Seguem-se duas referências:

> "É de se esperar que vetores, matrizes, transformações lineares, etc constituam a linguagem natural para tratar o Cálculo Diferencial pois, afinal de contas, este se baseia na idéia de aproximar, na vizinhança de cada ponto do seu domínio, uma função "arbitrária" por uma função linear (chamada sua derivada) e, a partir das propriedades desta (presumivelmente mais fáceis de constatar) obter informações sobre aquela".
>
> E. L. Lima

[1] E. L. Lima, *Curso de Análise, vol. 2* (3ª edição.) Coleção Projeto Euclides, IMPA, 1989.

A Álgebra Linear está presente em toda parte do estudo das funções reais de várias variáveis: na teoria das funções implícitas, nas integrais curvilíneas, na discussão dos pontos críticos (na qual as formas quadráticas desempenham o papel principal) e nas integrais de superfície, onde é reforçada por seu prolongamento natural, a Álgebra Multilinear.

[2] M. Hirsch e S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, 1974.

Nestes últimos 20 anos, o texto de Hirsch/Smale firmou-se como uma das principais referências para uma introdução moderna ao estudo das equações diferenciais ordinárias e uma preparação para a importante área dos sistemas dinâmicos. Ele pressupõe conhecimento de Análise a nível dos capítulos iniciais da referência 1 acima

e de Álgebra Linear a nível do presente livro. Mais de um terço do texto é dedicado às equações diferenciais lineares, o que leva os autores a desenvolver, de forma convincente, todo material de Álgebra Linear não tratado neste nosso livro. Na realidade, os autores afirmam que o estudo dos sistemas de equações diferenciais lineares com coeficientes constantes praticamente se identifica com um capítulo da Álgebra Linear.

Vejamos agora algumas referências de livros sobre Álgebra Linear.

Há poucas décadas eram raros, muito raros mesmo, os livros de Álgebra Linear destinados a estudantes de graduação. Hoje em dia há centenas deles, refletindo a enorme expansão do ensino desta disciplina aos mais variados cursos universitários. (Ver, a respeito, a citação de I. Kaplansky abaixo.)

Aqueles que mencionarei representam uma amostra, extremamente restrita, de três tipos: os livros onde aprendi a matéria, os que podem servir de leitura colateral e os que oferecem alternativas para estudos posteriores. A priori, aceito a acusação de parcialidade nas escolhas. Afinal, de gustibus et coloribus...

"It is desirable to have at hand not merely the formal operations with matrices, but also the (often neglected) interpretation of the matrices by linear transformations".

G. Birkhoff e S. MacLane

[3] Garrett Birkhoff e Saunders MacLane, *A Survey of Modern Algebra*. (Macmillan, 1941. Revised edition 1953.)

Birkhoff/MacLane é uma das mais bem sucedidas introduções à Álgebra Moderna já escritas. Trinta e oito por cento do livro é dedicado à Álgebra Linear. Embora a exposição seja feita a partir dos conceitos de espaço vetorial e transformação linear, a ênfase dominante é posta nas matrizes e suas "formas canônicas", especialmente matrizes de formas quadráticas. O estilo é ameno e ligeiro, bastante agradável. O leitor logo se acostumará com a notação vA, em vez de Av, usada pelos autores.

"That Hilbert space theory and elementary matrix theory are intimately associated came as a surprise to me and to many colleagues of my generation only after studying the two subjects separately. This is deplorable... I present this little book in an attempt to remedy the situation."

Paul R. Halmos

[4] Paul R. Halmos, *Finite Dimensional Vector Spaces*. (Princeton Univ. Press 1942. Revised edition: Van Nostrand, 1958. Tradução brasileira: Editora Campus, 1978.)

O livro de Halmos é outro best-seller. Nele, o autor conversa com o leitor e procura motivar, com analogias, os conceitos e as proposições, tudo isso feito com clareza e coerência lógica. Há uma grande preocupação em dar definições e demonstrações sem utilizar coordenadas. Isto é feito com o propósito admitido de preparar a caminho para os espaços de dimensão infinita, estudados em Análise Funcional. Só que muitas vezes esse purismo se torna artificial. Além disso, com vistas a diversas aplicações, a familiaridade do estudante com bases e coordenadas seria de grande utilidade.

Os livros acima, na ordem citada, me serviram de cartilhas de Álgebra Linear, como a tantos estudantes de várias gerações por todo o mundo. Eles contêm visões complementares sobre o assunto e sobre a maneira de ensiná-lo.

"Dealing with vector spaces in the abstract also saves effort. The general theory of vector spaces includes not only the vectors and matrices discussed in Chapters 1 and 2 but also sets of real- and complex-valued functions of a real variable and other more exotic mathematical objects. A fact which has been proved once and for all in the general theory applies to a wide range of particular cases. That is why it is worth investing some intellectual effort in understanding abstract linear algebra".

D. H. Griffel

[5] D. H. Griffel, *Linear Algebra: A First Course and Applications* (2 vols). (Ellis Horwood Limited, 1989.)

O livro de Griffel é escrito para estudantes não-matemáticos que deverão usar Álgebra Linear em suas carreiras. Ele emprega predo-

minantemente matrizes e $\mathcal{R}^n$ em vez de transformações lineares e espaços vetoriais porém, conforme a advertência acima citada, ocasionalmente se rende à conveniência de adotar uma atitude mais adequada. Trata-se de um livro de grande simplicidade, muito claro e bem organizado, com um sabor nitidamente "aplicado". Contém alguns erros matemáticos engraçados (como, por exemplo, "provar" que toda matriz anti-simétrica tem determinante zero ou afirmar que, pelo Teorema de Cayley-Hamilton, a exponencial de qualquer matriz reduz-se a um polinômio nessa matriz). Tais erros são poucos, não interferem no mérito geral e deverão ser corrigidos em próximas edições. São até instrutivos, inclusive para mostrar ao leitor o que pode esperar no futuro ao ler alguns livros de Matemática Aplicada.

> "The solution to each problem immediately follows the statement of the problem. However, you may wish to try to solve the problem yourself before reading the given solution. In fact, even after reading the solution, you should try to resolve the problem without consulting the text. Used thus, '3000 Solved Problems in Linear Algebra' can serve as a supplement to any course in linear algebra, or even as an independent refresher course".
>
> S. Lipschutz

[6] Seymour Lipschutz, *Schaum's solved problems series: 3000 solved problems in Linear Algebra*. McGraw-Hill Book Company, 1989.

A bem-sucedida série Schaum de livros de problemas se baseia numa idéia tipo ovo-de-Colombo: em vez de disputar sua adoção contra tantos e tão fortes competidores, esses livros-texto se disfarçam em coleções de problemas e assim convivem pacificamente com seus rivais, sendo adquiridos pelos estudantes, mesmo quando não recomendados pelos professores, como fontes suplementares de exercícios. De um modo geral (e isto se aplica ao livro de Lipschutz) eles contêm uma boa lista de problemas rotineiros, que não exigem grandes rasgos de imaginação. Mas, principalmente porque são acompanhados de solução, esses exercícios são uma ajuda valiosa para os alunos que necessitam um esforço adicional a fim de acompanhar o curso. Existe um livro análogo, também de Lipschutz, chamado "Linear Algebra", que foi traduzido para o português e publicado pela McGraw-Hill do Brasil, em 1972.

Os livros [5]. e [6]. acima se enquadram na categoria de leitura colateral. Seguem-se três referências a livros que constituem opções para a continuação deste texto.

"Matrix theory can be studied with no mention of linear spaces and most of the results in this book are of such a nature. However, the introduction of linear spaces and the role of matrices in defining or representing linear transformations on such spaces add considerably to our insight. Most important, perhaps, the notions of linear spaces and linear transformations give a geometrical basis to matrix theory, which aids both in understanding, as well as in suggesting proofs and new results.

J. Ortega

[7] James Ortega, *Matrix Theory. A Second Course.* Plenum Press, 1987.

A economia de pensamento e notação, bem como a riqueza imaginativa que provém do uso da linguagem geométrica resultam do emprego judicioso das noções de espaço vetorial e transformação linear. Ortega tira grande proveito desse ponto de vista intrínseco e consegue escrever um livro que, em meras 250 páginas, faz uma revisão dos princípios básicos da Álgebra Linear e desenvolve, com notável eficiência, uma exposição sobre tópicos avançados da álgebra das matrizes, que pode ser útil tanto para o matemático puro como para aqueles que se interessam de modo inteligente pelo cálculo numérico matricial.

"Linear Algebra, like motherhood, has become a sacred cow. It is taught everywhere; it is reaching down into the high schools; it is jostling calculus for the right to be taught first".

I. Kaplansky

[8] Irving Kaplansky, *Linear Algebra and Geometry.* Chelsea, 1969.

Kaplansky é um consagrado expositor. Na Universidade de Chicago (onde era colega de MacLane e Halmos) suas aulas eram famosas pela elegância das demonstrações e pelo notável poder de síntese. Estas qualidades estão presentes neste livro. Nele, o autor oferece uma alternativa para um segundo curso de Álgebra Linear, como fundamento básico da Geometria, esta última vista em toda a sua generalidade.

"As a very simple example the reader should think of the principal axis theorem (spectral theorem) for $\mathcal{R}^n$ which says that given a self-adjoint transformation, one can choose an orthonormal basis in $\mathcal{R}^n$ so that the matrix of that transformation in that basis is diagonal. That is, if one chooses the right isomorphic copy of $\mathcal{R}^n$ (change of basis) then the operator becomes especially simple. As the reader will see, this example is the first note of a rather long simphony".

M. Reed e B. Simon

[9] M. Reed/B. Simon, *Methods of Mathematical Physics, vol. I: Functional Analysis.* (Revised Edition, Academic Press, 1980.)

Ao mencionar o bem conhecido texto de Reed/Simon, minha intenção é apresentar um livro de Análise Funcional, uma área da Matemática onde tudo se passa dentro de espaços vetoriais. Há muitos bons livros sobre este assunto. (Um exemplo à mão é o excelente "Operadores Auto-Adjuntos e Equações Diferenciais Parciais", de Javier Thayer, publicado no Projeto Euclides do IMPA.) A escolha de Reed/Simon se deve não apenas às suas boas qualidades intrínsecas como também ao fato de que exibe a Análise Funcional (portanto os espaços vetoriais) como porta de entrada para a Física Matemática.

Uma palavra sobre pré-requisitos: o livro de Ortega está ao alcance imediato de quem leu o presente texto. Kaplansky requer um conhecimento elementar de corpos, a nível de um curso introdutório de Álgebra. Reed/Simon (ou qualquer outro livro de Análise Funcional) pressupõe noções básicas de Análise, Teoria da Integral e Equações Diferenciais.

[10] Ralph Costa Teixeira, *Álgebra Linear, exercícios e soluções*. (Coleção Matemática Universitária, IMPA, 2009.)

O livro de Ralph Costa Teixeira contém as soluções dos 594 exercícios propostos no presente texto. Na verdade, o total é bem maior do que seiscentos, pois vários desses exercícios são múltiplos. Além das soluções, todas completas e elegantemente apresentadas, cada capítulo tem início com a revisão dos conceitos a serem tratados, a discussão de simples exemplos adicionais e o destaque de algumas proposições referentes ao tema estudado.

Lista de Símbolos

Índice Remissivo

Índice Remissivo

O autor:

Elon Lages Lima é Pesquisador Emérito do Instituto Nacional de Matemática Pura e Aplicada (IMPA), Doutor Honoris Causa pela Universidade Federal de Alagoas, Universidade Federal do Amazonas e Universidad Nacional de Ingenieria del Perú, professor Honoris Causa da Universidade Federal do Ceará, da Universidade Federal da Bahia, da Universidade Estadual de Campinas e da Pontificia Universidad Católica del Perú, além de membro titular da Academia Brasileira de Ciências e da Academy of Sciences for the Developing World.

É autor de vários livros de Topologia, Análise, Álgebra e Matemática Elementar, dois dos quais são ganhadores do Prêmio Jabuti.

A coleção

A coleção Matemática Universitária é uma série de livros escritos por matemáticos competentes e com grande experiência didática, a fim de servirem de textos para cursos em nível de graduação nas universidades brasileiras.

Os livros da coleção contêm exposições objetivas e bem organizadas, acompanhadas de exercícios selecionados.

O livro

É uma introdução à Álgebra Linear, escrita para leitores que não necessitam possuir conhecimentos anteriores sobre o assunto. A apresentação busca um equilíbrio entre a organização teórica da matéria e seus aspectos concretos (ou práticos), como a eliminação gaussiana, o método de Lagrange para diagonalização de formas quadráticas ou as várias decomposições de matrizes.

Alguns tópicos, como valores singulares, pseudo-inversa, equações a diferenças finitas e outros, geralmente não discutidos em livros elementares como este, foram incluídos em vista de sua importância nas aplicações e também por servirem de atraentes ilustrações da teoria.

INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA